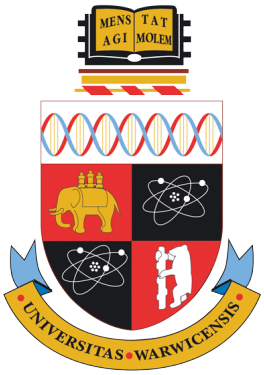


Combined Molecular Dynamics and Experimental Approach to Understand the Impact of the Oxetane and Azetidine Modification on Peptides

Eleanor Sophie Jayawant

Thesis submitted to the University of Warwick towards award of the degree of Doctor of Philosophy

Department of Chemistry

March 2021

Contents

List of Tables

List of Figures

List of Abbreviations

Alx	Oxetane-modified alanine (three letter code)
A_{ox}	Oxetane-modified alanine (single letter code)
AMP	Antimicrobial peptide
ANOVA	Analysis of variance
BBB	Blood-brain barrier
CaMHB	Cation-adjusted Mueller-Hinton broth
CD	Circular dichroism
CFU	Colony-forming unit
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at Harvard Macromolecular Mechanics
COSY	Correlation spectroscopy
CSA	Cyclosporin A
DEPBT	3-(diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one
DMF	Dimethylformamide
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
ffTK	Force Field Toolkit
G_{az}	Azetidine-modified glycine (single letter code)
Glx	Oxetane-modified glycine (three letter code)
Glz	Azetidine-modified glycine (three letter code)
G_{ox}	Oxetane-modified glycine (single letter code)
GROMACS	Groningen Machine for Chemical Simulations
H_{ox}	Hydrogen atom of oxetane ring
HSQC	Heteronuclear single quantum coherence
HPLC	High performance liquid chromatography
L-J	Lennard-Jones
MD	Molecular dynamics
MBC	Minimum bactericidal concentration
MIC	Minimum inhibitory concentration

NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser enhancement spectroscopy
PEG	Polyethylene glycol
PPI	Protein-protein interaction
QM	Quantum mechanics
ROE	Rotating-frame Overhauser effect
ROESY	Rotating-frame Overhauser enhancement spectroscopy
RNA	Ribonucleic acid
SEM	Scanning electron microscopy
SMD	Streered molecular dynamics
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TMS	Tetramethylsilane
TOCSY	Total correlation spectroscopy
U-B	Urey-Bradley
VMD	Visual Molecular Dynamics
WHO	World Health Organisation

Acknowledgements

There are many people that I need to thank for their help in both writing this thesis and completing my PhD. From an academic standpoint, I owe huge thanks to my brilliant supervisors, Dr Ann Dixon and Dr Becky Notman, for pushing me to work to the best of my capabilities and for making sure I made it through my PhD with my mental health intact. Many thanks to my advisory panel, Prof Józef Lewandowski and particularly Dr Gabriele Sosso, who have been a consistent source of feedback and support, and in the case of Gabriele going so far as to supervise me whilst Becky was away on maternity leave. I also must thank Prof Mike Shipman for welcoming me into his group and providing regular feedback on my work. Dr Ivan Prokes and John Moat have taught me technical skills which were essential throughout my PhD. There are many other people who have helped me, by letting me bounce ideas off of them: Chris, Gareth, Christine, Rhiannon and Ciaran have all been open and ready to listen when I have approached them with problems, both work-associated and personal. Stefan, George and Ina from Shipman group have tolerated my stupid questions and tirelessly prepared peptides for me. I should thank my 12:30 Lunch Crew, for providing a work-free dedicated lunch time (which was a much-needed time for blowing off steam), as well as the occasional cheeky pub time. Punit also assisted in the preparation and formatting of my graphs by providing a script, which I am very grateful for.

My parents have provided me with huge amounts of love and support (both emotional and financial), so many thanks to them, along with the rest of my family and friends. When we first moved to Warwickshire three years ago we didn't know anyone, and I am so grateful for the wonderful support network we've built up here. There are too many of you to list here, but I treasure our friendship and I hope you know that you've helped keep me sane during the wedding planning, house moving, poor mental health and literal pandemic that have sometimes made this PhD seem impossible. Finally, this thesis wouldn't exist without the support of my husband, James. James pushes me to do my best every day, and I am eternally grateful for his endless love and proofreading.

Declaration

The work presented here in this thesis is original, and was conducted by the author, except where otherwise stated, under the supervision of Dr Ann M. Dixon and Dr Rebecca Notman at the University of Warwick between October 2017 and March 2021.

Several of the peptides used in this work were prepared by Dr Stefan Roesner, Dr Ina Wilkening, Dr Jonathan Beadle, Dr George Saunders and Dr Sam Spring, University of Warwick, UK. The circular dichroism data presented in Chapter 5 was collected and processed by Dr Jonathan Beadle, University of Warwick.

This work has not previously been presented to be considered for another degree.

All sources of information have been acknowledged in the form of references.

Funding was provided by the Leverhulme Trust.

At the time of submission, parts of this thesis have been published as detailed below:

Jayawant, E. S., Beadle, J. D., Wilkening, I., Raubo, P., Shipman, M., Notman, R., & Dixon, A. M. (2020) Impact of oxetane incorporation on the structure and stability of alpha-helical peptides. *Phys. Chem. Chem. Phys.*, *22*(43), 25075–25083.

Roesner, S., Saunders, G. J., Wilkening, I., Jayawant, E., Geden, J. V., Kerby, P., Dixon, A. M., Notman, R., & Shipman, M. (2019) Macrocyclisation of small peptides enabled by oxetane incorporation. *Chem. Sci.*, *10*(8), 2465–2472.

Abstract

Peptide-based drugs are promising as therapeutics, but they are limited by unfavourable pharmacokinetic properties. Modifying or cyclising the peptide backbone increases their biological activity in many cases. However, cyclic peptides remain challenging to make, and no single backbone modification is universally tolerated. Both oxetanes and azetidines have previously been used in medicinal chemistry, and previous work suggests that oxetane modification of the peptide backbone may bring the termini in close proximity, enhancing efficiency of cyclisation. In order to realise the potential of oxetane and azetidine modification in medicinal chemistry, an investigation into the effects of these modifications on the backbone of peptides is needed.

In this work, linear and cyclic oxetane-modified peptides were characterised using CD, NMR spectroscopy and molecular dynamics (MD) simulations. In linear peptides, oxetane modification induces the formation of a turn, as evidenced by the presence of medium-range NOEs observed in NMR experiments. In cyclic peptides, MD simulations reveal that the oxetane-modified peptide is more rigid and the structure is stabilised by the formation of a new hydrogen bond that is absent in the parent peptide. The effect of oxetane modification on α-helicity was also explored. CD spectroscopy revealed that oxetane modification drastically reduces the helical content of a well-characterised sequence. Furthermore, steered MD simulations revealed the molecular determinants of this destabilisation—that oxetane modification introduces a kink in the helical axis, which disrupts the dihedral angles and hydrogen bonding pattern in both directions.

NMR experiments showed that, like oxetane, azetidine induces a turn in a short linear peptide. Finally, the impact of azetidine modification on the activity of antimicrobial peptides (AMPs) was assessed using microbiological assays. Following confirmation of activity, a known AMP sequence was modified and/or cyclised, and the activity was measured. It was found that azetidine modification increases the activity of this AMP without increasing its toxicity.

Overall, this work may help direct future design of peptide-based drugs.

Chapter 1

Introduction

1.1 Statement of the Problem

Peptide-based drugs have the potential to combine the advantages of traditional small molecule drugs and protein-based biologics, but they often have disadvantageous physicochemical properties. Chemical modification of peptides for improved bioavailability and metabolic stability is a critical aspect of work in the development of peptide-based drugs.[1] Introduction of non-natural backbone linkages such as thioamides, azapeptides (where the Cα is replaced with a nitrogen) and poly-N-substituted glycines (peptoids), has led to the increased stability or enhanced bioavailability of a wide range of peptides.[2–6] However, backbone modification has also been shown to fundamentally change the structural properties of polypeptides, and no single modification is universally tolerated. For this reason, it is important to understand of the impact of available modifications for future development of peptide-based medicines.

In this work, we explore the impact of oxetane modification on peptide structure using a combination of biophysical and computational techniques, building on previous work by Shipman and co-workers.[7–10] Synthesis of oxetane-modified peptides can be challenging, and more recently Shipman and co-workers have been exploring the synthesis and structure of azetidine-modified peptides. The impact of azetidine modification is also described in this work, using a mixture of biophysical and microbiological techniques.

1.2 Drugs

1.2.1 Traditional Drugs

Drug molecules tend to fall into one of two categories: traditional small molecule drugs and larger, protein-based biologics. Many small molecule drugs meet the criteria specified by Lipinski's well-known 'rule-of-five',[11] a set of *in silico* guidelines which note a preference for a molecular weight of <500 Da to favour oral bioavailability.[12] The 'rule-of-five'[11] states that an orally active drug in general has no more than one violation of the following:

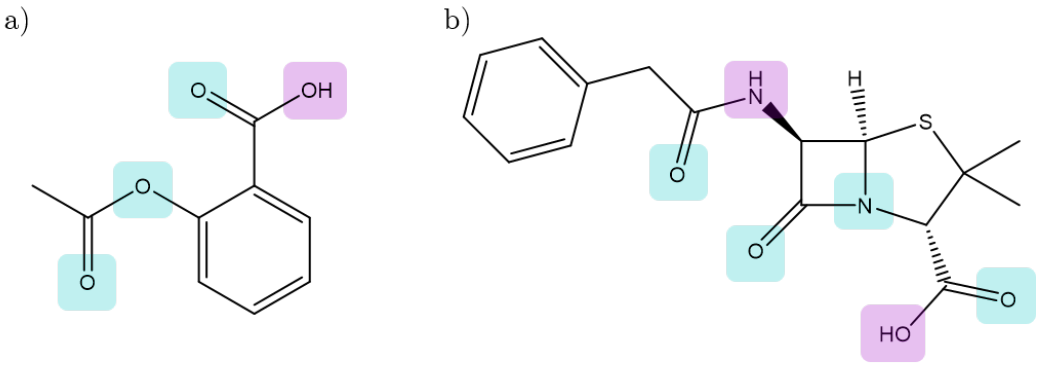
- No more than 5 hydrogen bond donors (expressed as the sum of OHs and NHs).
- No more than 10 hydrogen bond acceptors (expressed as the sum of Ns and Os).
- A molecular weight of less than 500 Da.
- An octanol-water partition co-efficient (Log P) of less than 5.

These physicochemical parameters are important, as they determine the ability of a drug molecule to cross cell membranes in order to reach the systemic circulation. Smaller molecules can more readily fill in the transient water pores formed in biological membranes. Similarly, drugs with low capacity for hydrogen bonding and a non-polar surface area can cross the hydrophobic phospholipid membrane bilayer more easily.[13]

As a result, small molecule drugs (such as aspirin, Fig. 1.1a, and penicillin, Fig. 1.1b) are less than 500 Da and are readily orally bioavailable. This is highly advantageous for a drug molecule—oral delivery is the preferred method for drug administration as it results in the highest rate of patient compliance.[14]

1.2.2 Biologic Drugs

Biologics are treatments derived from living organisms, such as hormones, antibodies, vaccines and antitoxins. The use of protein-based therapeutics is not a new concept, with treatments for diabetes mellitus using insulin derived from the pancreases of pigs available as early as the 1920s,[15] and this field rapidly expanded in the latter part of the 20th century thanks to advances in the fields of molecular biology, genomics and proteomics.[16] For example, adalimumab (sold under the brand name Humira), a recombinant monoclonal antibody drug used to treat inflammatory bowel disease and other immune conditions, has become one of the largest blockbuster drugs of all time.[17]



aspirin
180 Da

penicillin
334 Da

Figure 1.1: Structures of (a) aspirin, and (b) penicillin, two well-known examples of small molecule drugs. Hydrogen bond acceptors are indicated in blue and hydrogen bond donors are indicated in purple. Aspirin and penicillin do not violate any of the principles outlined by the rule-of-five.

Protein-based biologics disobey all of Lipinski's rule-of-five:[12] they are large (more than 5000 Da in size); and cannot be taken orally without being destroyed in the gastrointestinal tract. As a result, they are not orally bioavailable, and typically must be administered by injection. Additionally, their potency is difficult to quantify in comparison to small molecule drugs.[18] However, biologics have several advantages: biologic drugs are highly effective therapeutics in treating certain diseases, and they have fewer off-target effects compared to traditional small molecule drugs.[16]

There is a large 'gap' in molecular weight between the two categories of small molecules and biologic drugs, suggesting that there is the potential to exploit molecules between 500–5000 Da (Fig. 1.2). Such molecules may combine the advantages of small and large drugs.[16]

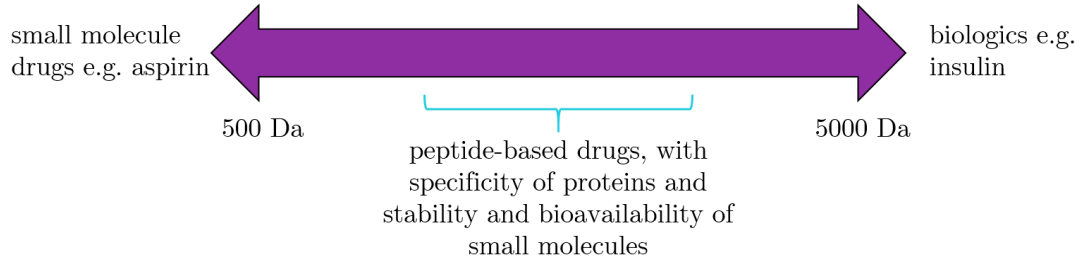


Figure 1.2: Schematic representation of the molecular weight 'gap' between traditional small molecule drugs, which are less than 500 Da, and larger biologic drugs, which are more than 5000 Da. Peptides may fill this 'gap', and can potentially combine the advantages of both large and small drugs. Reproduced from Craik et al.[16]

1.3 Peptides as Drugs

1.3.1 Peptides

Peptide-based drugs may be able to fill this molecular weight 'gap'. Peptides are biomolecules that are composed of 2–50 amino acids joined by peptide bonds (also called amide bonds, Fig. 1.3). Peptide bonds are formed by a condensation reaction between the carboxylic acid end of one amino acid and the amine group of the next. Under normal physiological conditions, peptides are zwitterionic, and the ends of a peptide are referred to as the N- or C-terminus based on their end groups (NH_3^+ or COO^- respectively).

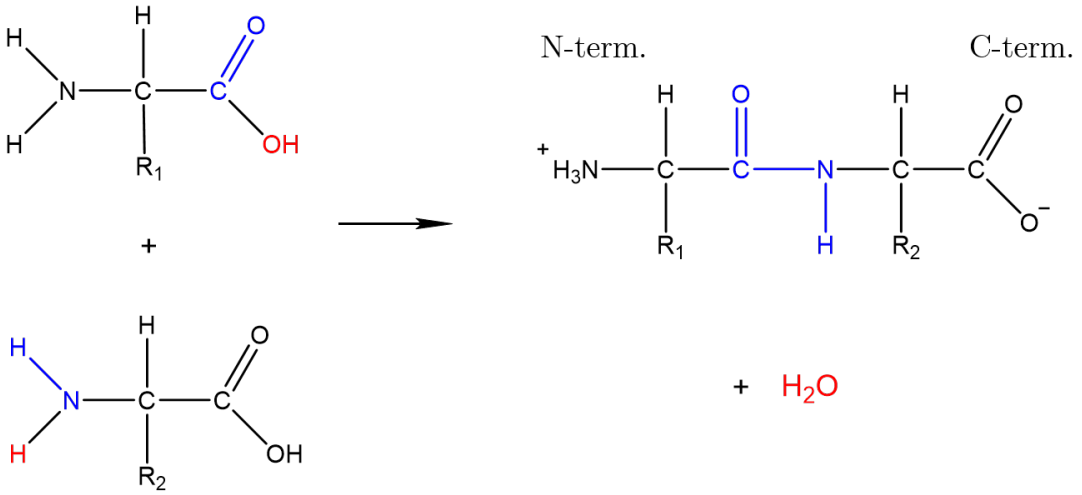


Figure 1.3: Peptide bond formation by condensation. Amino acid subunits (R = sidechain) come together to form a dipeptide. Under normal physiological conditions, both the N- and C-terminus are charged (zwitterionic).

The structure and function of peptides is highly dependent on the amino acid sequence—different amino acids have different properties imparted by the size, shape and chemical properties of their sidechains (R groups). There are 20 standard amino acids, ranging from 75–204 Da in size. The smallest amino acid, glycine, has a sidechain consisting of a single hydrogen atom.

Although peptides (and their larger counterparts, proteins) are constructed from simple building blocks, oligomerisation results in complex structures that play vital roles in almost all biological processes. Certain residues are more common in secondary structural elements such as α-helices and β-sheets, which are important for biological function, due to the different chemical properties of the sidechains—for example, a peptide with a high proportion of hydrophobic residues such as leucine or alanine might form stable helices in certain solvents.[19] Peptides have many critical roles in human physiology, acting as hormones, growth factors,

neurotransmitters, and more.[20] As a result, the therapeutic application of peptides and proteins is of great interest.[21–23]

Peptides remain promising as drug molecules, and from the 1960s to the turn of the millennium they were often considered the drugs of the future.[24] Peptides represent a significant portion of the biopharmaceutical industry, with over 50 peptide-based drugs on the market already, with hundreds in clinical trials and preclinical development.[25] Some of the advantages and disadvantages of peptide-based drugs are summarised below (Table 1.1). Peptide-based drugs can be readily synthesised and have fewer toxic side effects compared to small molecule drugs.[16]

Table 1.1: Potential advantages and disadvantages of peptide-based drugs. Reproduced from Craik et al.[16]

Advantages	Disadvantages
High potency	Poor metabolic stability
High selectivity	Poor membrane permeability
Broad range of targets	Poor oral bioavailability
Potentially lower toxicity than traditional small molecule drugs	High production costs
Low accumulation in tissues	Rapid clearance
High chemical and biological diversity	May have poor solubility
Discoverable at peptide and/or nucleic acid levels	

Despite their advantages, there are still significant challenges with regards to the use of peptide-based drugs. Peptides typically have poor oral bioavailability—they are readily degraded by proteases and have poor permeability across the intestinal membrane.[26] They are also prone to agglutination (clumping) and opsonisation (targeted destruction by phagocytes).[14] Furthermore, compared to cheap and small molecules, the costs of developing a new peptide-based drug are still high, especially for longer peptides.[27] Naturally-occurring peptides are often not directly suitable for use as therapeutics, as they have poor chemical and physical stability, and a short half-life in plasma.[21]

Perhaps the best-known example of a peptide drug is cyclosporin A (CSA). It was originally isolated from the fungus *Tolypocladim inflatum*,[28] and is an immunosuppressant that has revolutionised human organ transplant therapy, consistently appearing in the World Health Organization's List of Essential Medicines. Unlike many peptide drugs, CSA has acceptable oral bioavailability, with values ranging from 10–60% depending on formulation.[29]

As an 11-residue peptide, CSA violates two of Lipinski's rules: it is larger than 500 Da, and has more than 10 hydrogen bond acceptors. However, it has some

distinct features (Fig. 1.4) that are thought to contribute to its impressive oral bioavailability. CSA contains a macrocyclic backbone, which reduces the conformational space sampled and can promote permeability.[29] Additionally, it contains seven N-methyl groups, which reduces the number of hydrogen bond donors. N-methylation is a desirable trait for a drug molecule, as the reduction in intermolecular hydrogen bonding increases lipophilicity and is strongly correlated with cell permeability of small molecule and peptide-based drugs.[14,29] Finally, it also contains two non-canonical amino acids (i.e. not one of the 20 proteinogenic amino acids), and a D-alanine residue.[28] Other peptide-based drugs such as somatostatin and encephalin have similar characteristics to CSA and also have improved oral absorption after cyclisation.[14]

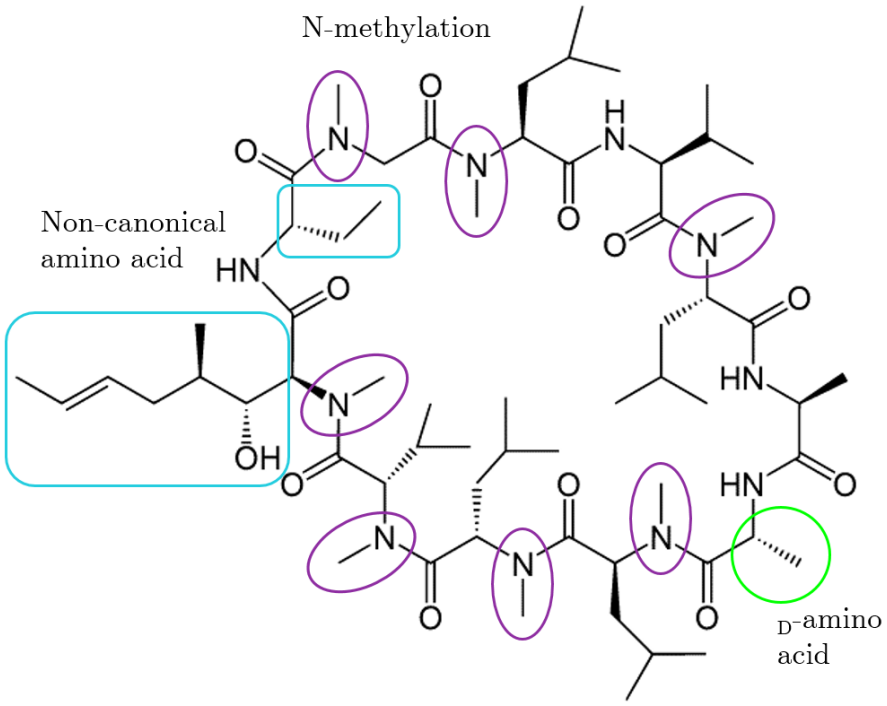


Figure 1.4: Structure of cyclosporin A, with key characteristics highlighted. It contains two residues with non-canonical sidechains (cyan), seven N-methylated residues (purple) and a D-amino acid (green). Reproduced from Ito et al.[28]

Because several of the innate physicochemical properties of peptides make them disadvantageous as drug molecules, interest has shifted from simpler naturally-occurring peptides (which are not resistant to proteases, agglutination or opsonisation) towards peptides which contain some of the features that make CSA so successful: modified peptides, peptidomimetics, and cyclic peptide-based molecules.[14]

1.3.2 Peptide Modifications

Peptide modifications are commonly used in the drug development process, with varying levels of success. PEGylation is one such modification option for some peptides which cannot be readily cyclised.[26] Polyethylene glycol (PEG) is a non-toxic amphipathic molecule (Fig. 1.5),[30] and direct PEGylation of peptides and proteins improves absorption and systemic stability, and increases circulation time. However, PEGylation also has some disadvantages—due to the size of PEG, steric hindrance may decrease the activity of the protein and may increase protein aggregation.[14]

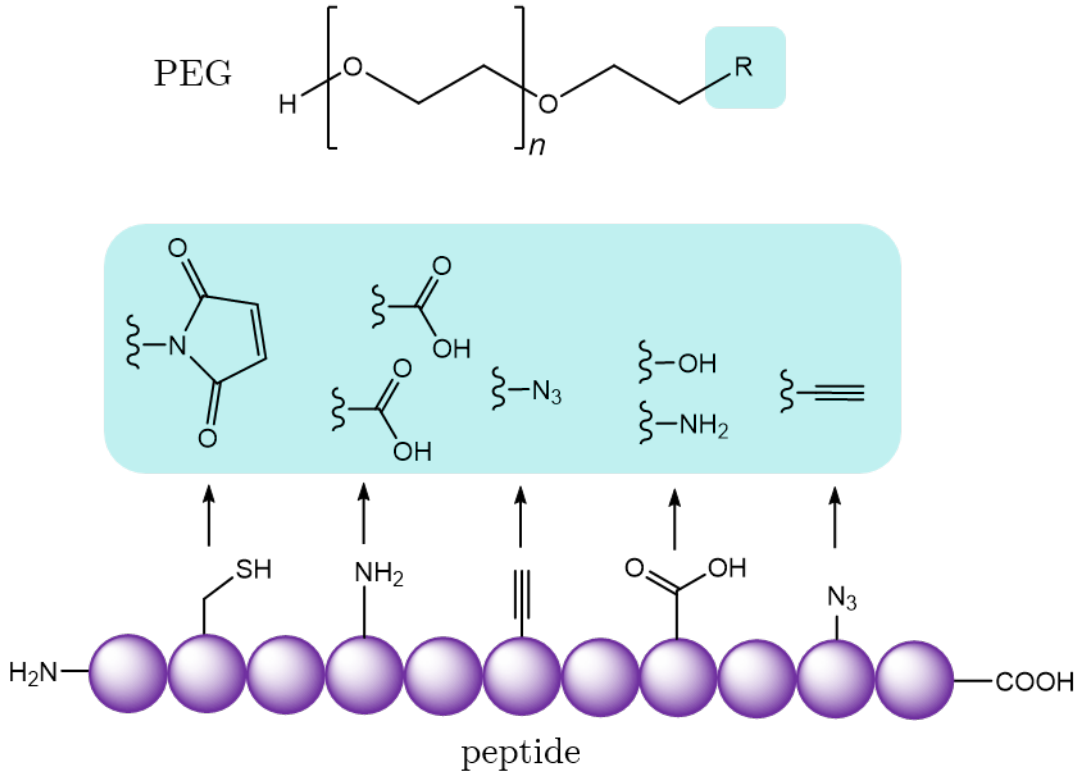
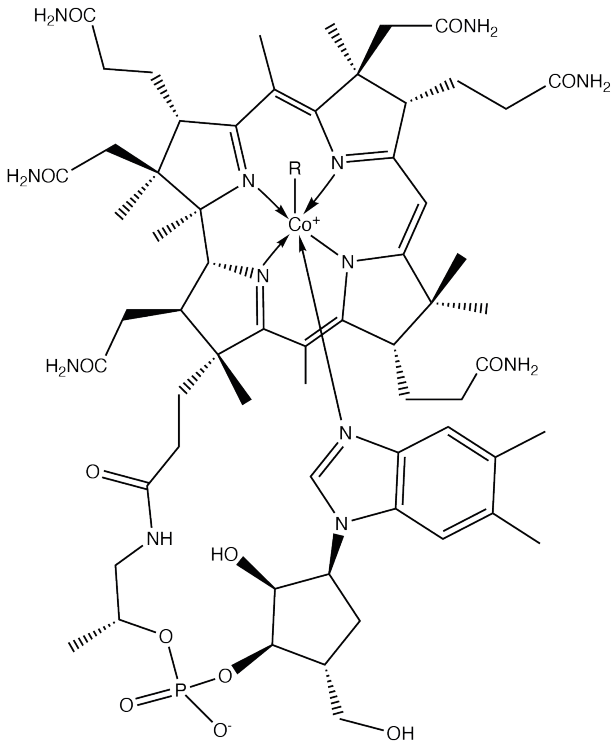


Figure 1.5: Structure of polyethylene glycol (PEG), and the chemical reactions used to introduce PEG units onto a peptide with commercially available building blocks. Reproduced from Erak et al.[30]

Direct conjugation to vitamin B12 (Fig. 1.6) is another promising method for delivery of peptide-based drugs. B12 is useful for oral drug delivery as it is both protected and absorbed from the gastrointestinal tract upon binding of haptocorrin, a protein produced by the salivary glands. Haptocorrin binds to B12 with high affinity in acidic conditions and protects it from acid hydrolysis.[31] In order to successfully use the B12 pathway, the peptide or protein must be coupled to B12 in a manner that neither molecule obstructs each other—B12 must still be recognised by haptocorrin, and the peptide or protein still needs to interact



vitamin B12

Figure 1.6: Chemical structure of vitamin B12. The core of the molecule is a corrin ring, which consists of 4 pyrrole-like subunits. The nitrogen of each pyrrole is coordinated to the central cobalt atom.

with its desired target. Successful conjugation of B12 to therapeutic peptides has the potential to increase oral bioavailability and plasma-residency time. However, challenges remain in delivery of a clinically relevant dose, as absorption may vary with total body stores of B12.[32]

Substitution of natural L-amino acids (Fig. 1.7a) with D-amino acids (Fig. 1.7b) is another strategy to enhance the oral bioavailability of peptides. Replacing certain residues with D-amino acids in a variety of peptides has been shown to reduce peptidic cleavage by trypsins and peptidases,[14] and to enhance binding affinity. For example, Chen and co-workers showed that substituting glycine residues in a bicyclic peptide inhibitor of the cancer-related protease urokinase-type plasminogen activator with D-amino acids improved inhibitory activity and stability.[33]

D-amino acids may also have enhanced ability to cross the blood-brain barrier (BBB), making them more effective at treating central nervous system diseases.[34] For example, Wei and co-workers demonstrated that a D-peptide ligand of nicotine acetylcholine receptors showed enhanced transcytosis efficiency across an artificial BBB compared with the parent L-peptide.[35] However, peptides containing D-amino acids are more expensive and more challenging to synthesise than their natural counterparts.

Other peptide modifications that can be used to change the physicochemical properties of peptide-based drugs include N-terminal derivatisation,[36] and lipidi-

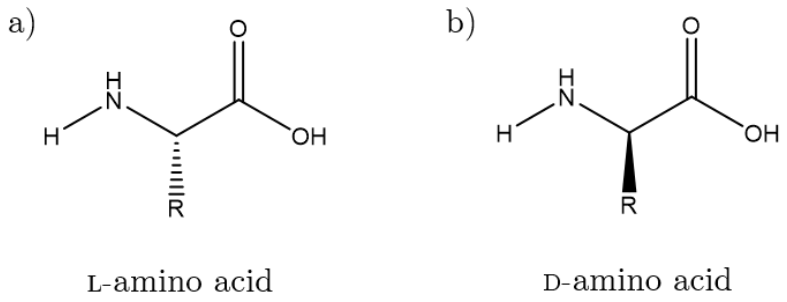


Figure 1.7: Structures of (a) L- and (b) D-amino acids (R = sidechain). The only difference between the structures is the chirality of the sidechain.

sation, which involves the covalent conjugation of a hydrophobic residue of the peptide to a hydrophobic compound and increases lipophilicity of the peptide.[37]

Several common peptidomimetic backbone modifications are summarised in Fig. 1.8, although this is not an exhaustive list and many others have been reported in reviews.[38–40]

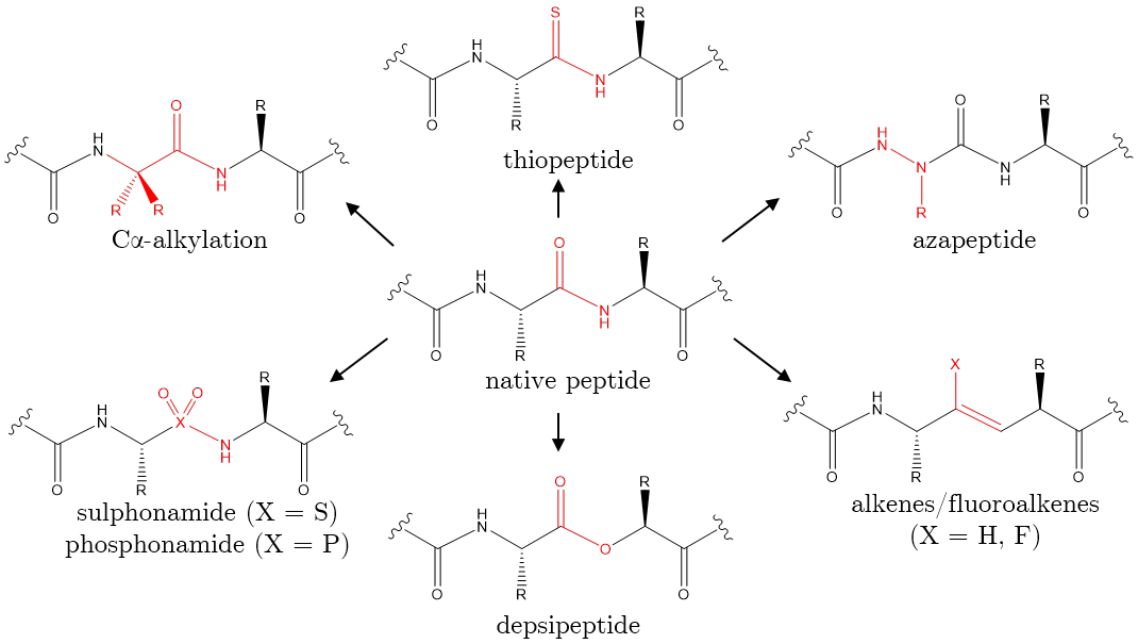


Figure 1.8: Examples of several common backbone modifications. R = amino acid sidechain. Adapted from Saunders.[41]

1.3.3 Cyclic Peptides

There are numerous examples of cyclic peptide therapeutics, with over 40 currently in the market, ranging from immunosuppressants such as CSA to antibiotics including vancomycin and polymyxin B to hormone analogues such as vasopressin.[42]

Cyclisation can occur in several different ways (Fig. 1.9), such as head-to-tail (N- to C-terminus) as for CSA (Fig. 1.4), or between the sidechains, as for lanthipeptides.[20]

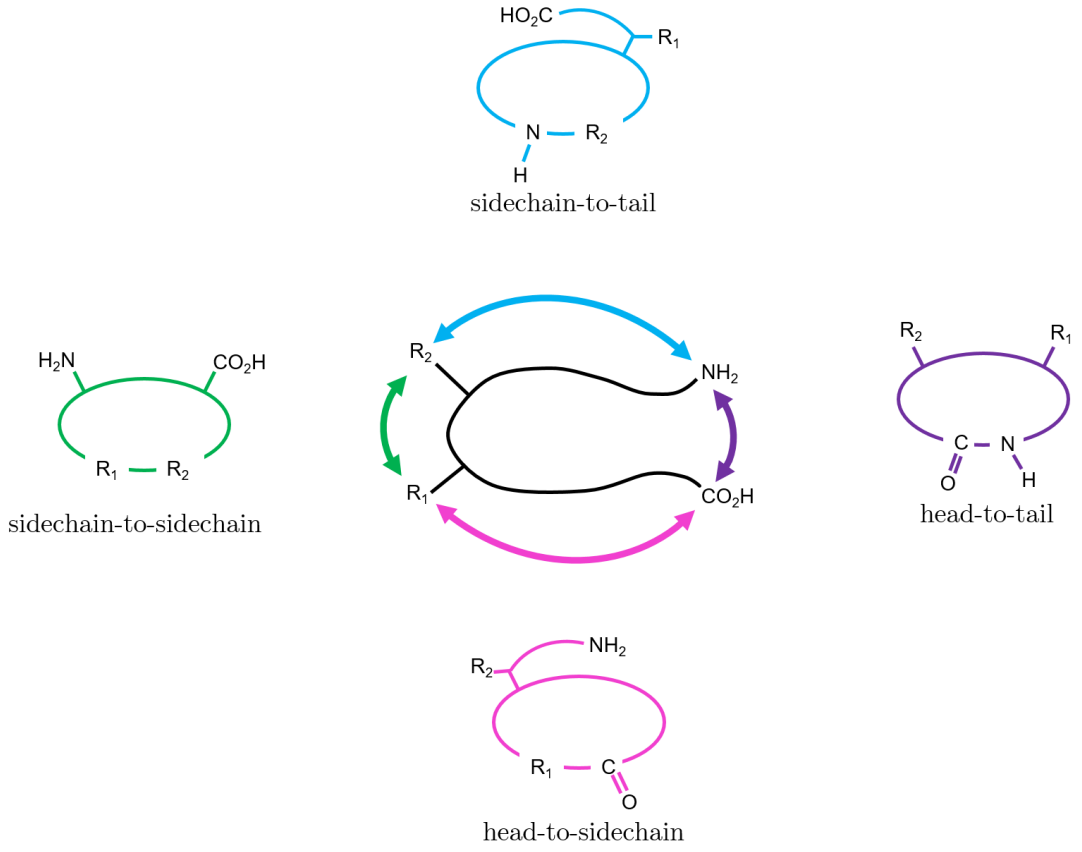


Figure 1.9: Four possible ways peptide cyclisation can occur: sidechain-to-sidechain (green), sidechain-to-tail (blue), head-to-tail (purple) and head-to-sidechain (pink). Arrows indicate where the new bond is formed to join the macrocycle. Reproduced from White and Yudin.[43]

Cyclic peptide backbones appear to be a valuable attribute when selecting a peptide-based drug, as it protects against proteolytic degradation and helps to bury polar groups in the interior of the molecule.[16] Due to the lack of C- and N-termini, head-to-tail cyclisation confers resistance to both endo- and exo-peptidases.[44]

Cyclisation also affects passive permeability, for several reasons. Firstly, where head-to-tail backbone cyclisation has occurred, the absence of free termini reduces potential interactions with the solvent, which improves permeability. Secondly, a cyclic peptide is more compact compared to its linear counterpart, which allows it to diffuse faster through a membrane due to its reduced collision profile in solution. Finally, in comparison to a linear peptide of the same size, a cyclic peptide is more rigid and samples a more restricted conformational space. If the

conformations that are sampled favour permeability, then the cyclic peptide spends more of its time in those conformations compared to their linear counterparts, and therefore are able to permeate the membrane more frequently.[29] In this way, a cyclic backbone can be thought of as a sort of 'conformational lock'.

The increase in rigidity also has been shown to affect biological activity. Cyclisation can lead to a loss of activity if the reduced flexibility fixes the compound in an inactive conformation,[45] but in many cases cyclisation has been shown to increase biological activity.[42,44,46–48] For example, cyclised variants of peptides derived from myelin basic protein were shown to be more stable and with higher *in vivo* efficacy compared to their linear counterparts.[46]

However, cyclic backbones can be difficult to select for in large scale pharmaceutical screening studies, since mass spectrometry-based proteomics studies are typically poor at sequencing cyclic peptides without first linearising them. Additionally, most natural cyclic peptides were thought to be synthesised through non-ribosomal biosynthetic pathways, and this makes screening at the nucleic acid level impossible due to a lack of transcripts, although over the last two decades the number of reported ribosomally-synthesised cyclic peptides is increasing.[16]

Preparation of cyclic peptides synthetically is also challenging, especially where the peptide is small (less than seven amino acids in length) and does not contain a turn-inducing element such as proline.[49,50] Although solution-phase peptide synthesis can be used to generate cyclic peptides, this method is demanding and frequently has poorer yields compared to solid-phase strategies. Solid-phase peptide synthesis involves an insoluble and activated solid polymer support (resin), to which amino acid building blocks are added in a stepwise fashion, resulting in the growth of the peptide chain. The peptide is then cleaved from the resin, and can be cyclised, and the sidechain protecting groups can be removed.[46]

In order for a linear precursor to undergo macrocyclisation, it must adopt a conformation in which the reactive groups are in close proximity. However, because *trans* peptide bonds (Fig. 1.10a) are generally more favourable than *cis* peptide bonds, it is not energetically favourable for the reacting groups to be near each other (especially when cyclising a peptide with 4–6 amino acids).[43] Rather than the desired product, oligomeric by-products often predominate (Fig. 1.10b), and due to the slow reaction rates C-terminal epimerisation can occur (Fig. 1.10c). C-terminal epimerisation results in the stereochemistry of one of the amino acid sidechains to be inverted (i.e. resulting in a D-amino acid when the desired product contains only L-amino acids).

There are many strategies to overcome the challenges associated with head-to-tail macrocyclisation, including the use of turn-inducing amino acids, backbone modifications, and performing the reaction in under dilute conditions, but there is

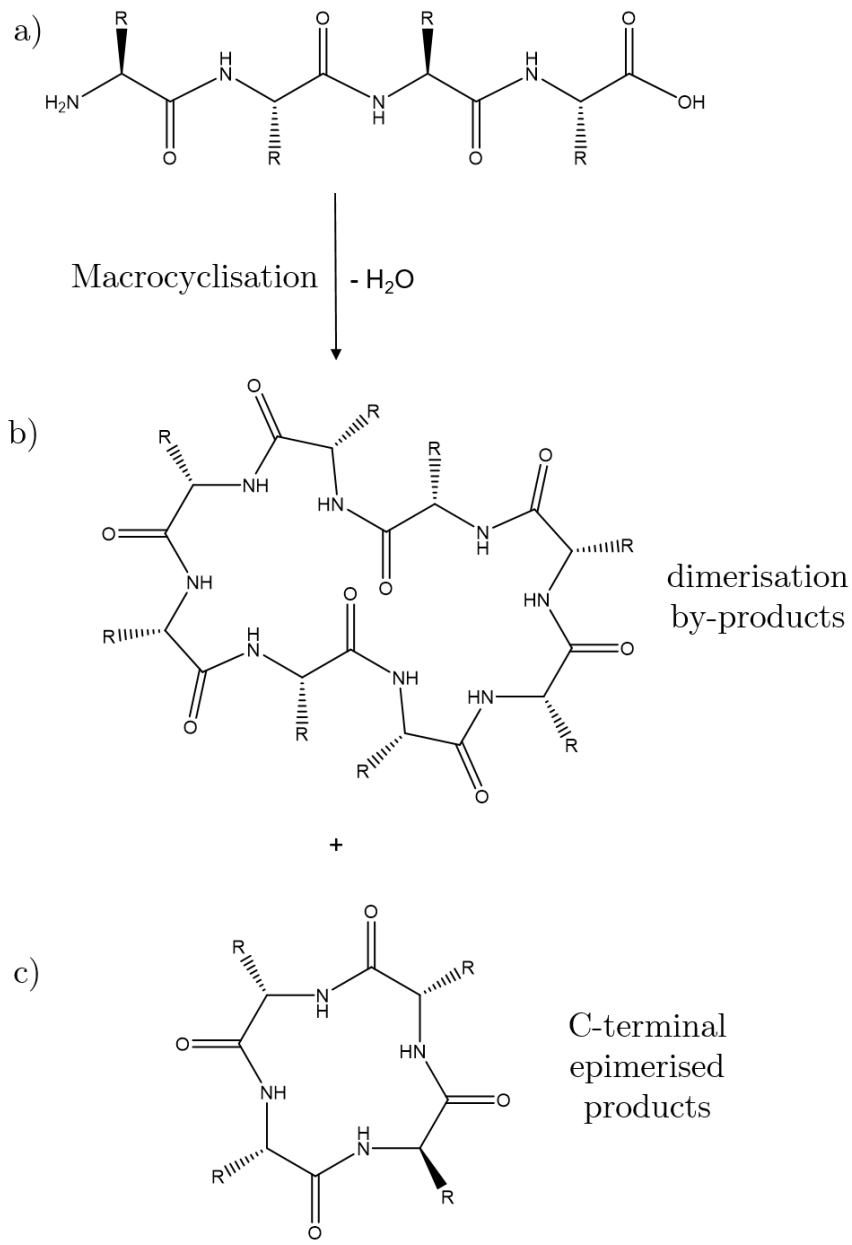


Figure 1.10: Limitations to macrocyclisation of (a) short peptide chains with *trans* amide bonds. Macrocyclisation can result in (b) oligomeric by-products, and/or (c) C-terminal epimerised products. Reproduced from Saunders.[41]

currently no general method for the preparation of small cyclic peptides.[43]

1.4 Oxetane

Introduction of oxetanes offers a promising route to producing cyclic peptides with enhanced efficiency. Oxetanes are four-membered heterocyclic compounds composed of three carbon atoms and one oxygen atom (Fig. 1.11). They have attracted considerable attention in medicinal chemistry due to their ability to

modulate the physicochemical properties of molecules.[51–53]

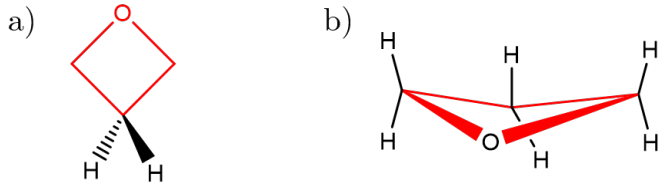
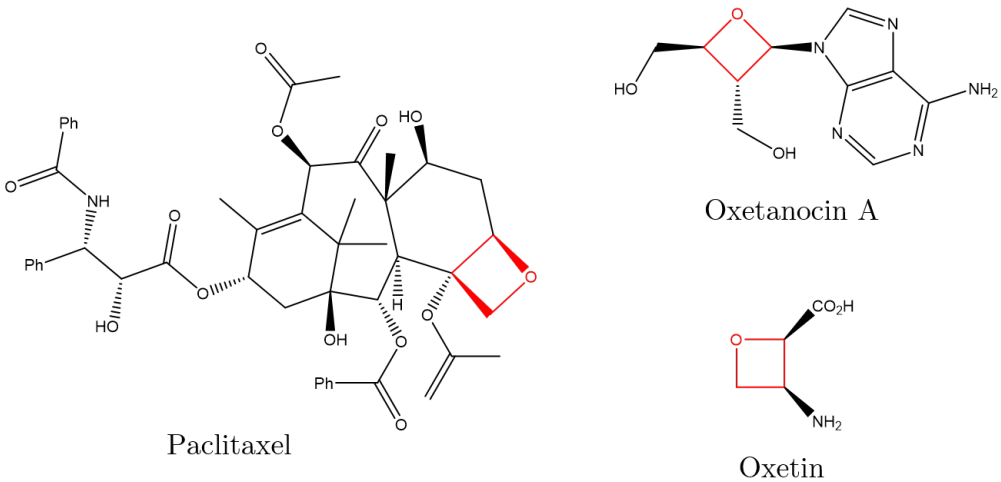


Figure 1.11: (a) Structure of the oxetane ring, and (b) ring puckering (10.7° at 90 K) of oxetane. Adapted from Bull et al.[53]

1.4.1 Oxetane in Nature

Although oxetane-containing natural products are rare, they have been found to exert interesting biological activities. These natural products range from fairly simple structures to complicated fused ring systems (Fig. 1.12). Perhaps the most well-known example is paclitaxel (marketed as Taxol), derived from the bark of Pacific Yew (*Taxus brevifolia*) and used in chemotherapy for many types of cancer.[54] In paclitaxel, the oxetane ring is proposed to act as a conformational lock, rigidifying the structure and allowing it to bind to microtubules.[53] Other examples include the nucleoside analogue oxetanocin A, which was isolated from soil bacteria and inhibits the *in vivo* replication of human immunodeficiency virus,[55] and the β-amino acid oxetin, which has antibacterial and herbicidal activity.[56]



Paclitaxel

Oxetanocin A

Oxetin

Figure 1.12: Oxetane-containing natural products. In each case, the oxetane is highlighted in red. Adapted from Bull et al.[53]

1.4.2 Oxetane in Medicinal Chemistry

Oxetane has previously been used as an isosteric replacement for *gem*-dimethyl and carbonyl groups in medicinal chemistry.[51,57,58]

gem-Dimethyl groups are often used in medicinal chemistry to block metabolic attack of methylene groups. However, replacement of the hydrogen atoms with methyl groups increases lipophilicity, which in turn may adversely affect its pharmacological properties.[58] Additionally, the *gem*-dimethyl group itself is at risk of metabolic attack. Previous work by Carreira and co-workers suggests that the oxetane ring can be used as a polar replacement of the *gem*-dimethyl group, as they have a similar molecular volume and oxetane reduces lipophilicity and metabolic vulnerability (Fig. 1.13).[51,58]

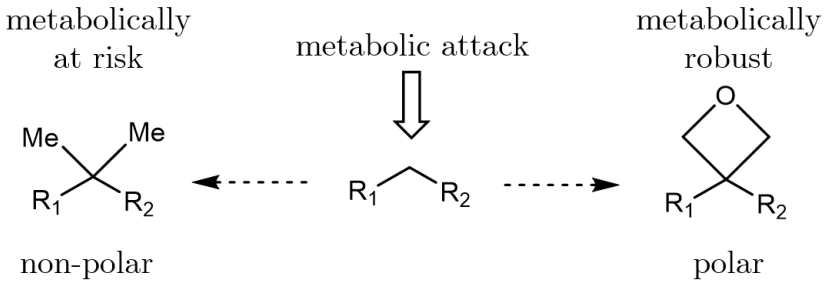


Figure 1.13: Comparison of *gem*-dimethyl and oxetane modules. The oxetane is metabolically robust compared to the *gem*-dimethyl unit. Reproduced from Wuitschik et al.[58]

Oxetane can be a valuable isosteric replacement for a carbonyl group due to their similar dipoles, hydrogen bonding capabilities and lone pair electron arrangement, although the oxetane has a larger length and volume (Fig. 1.14), which could alter pharmacokinetic properties such as binding. Small drug-like molecules in which the carbonyl group is replaced with an oxetane show low acidity and high permeability compared to unmodified compounds, providing advantages in the context of central nervous system drugs.[59] This is because a carboxylate COO^- is stabilised by resonance (Fig. 1.15a), resulting in a more stable conjugate base and lower pK_a compared to an alkoxide O− next to an oxetane ring (Fig. 1.15b. This reduction in acidity makes the compound less polar and more lipophilic, enhancing permeability. Additionally, while carbonyl groups are vulnerable to enzymatic attack and epimerisation of stereogenic centres, oxetane derivatives are stable to both of these concerns.[53]

An example of carbonyl to oxetane substitution is provided in work carried out by Carreira and co-workers, in which the drugs thalidomide and lenalidomide were modified.[60] Thalidomide is a sedative which was marketed for morning sickness during pregnancy. It was withdrawn from use after it was found to be teratogenic,

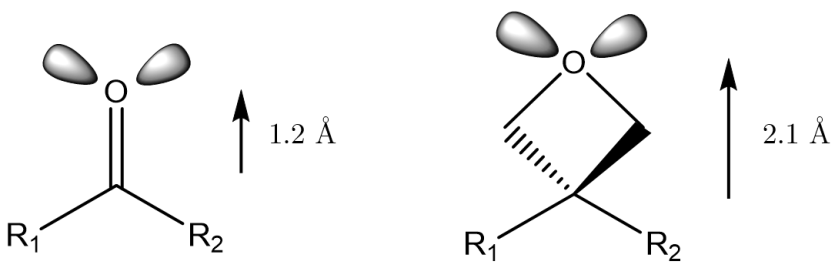


Figure 1.14: Comparison of carbonyl group with oxetane ring. The two functional groups show similar lone pair arrangement but the distance between the carbon and oxygen atoms in the oxetane is considerably larger than in the carbonyl. Reproduced from Bull et al.[53]

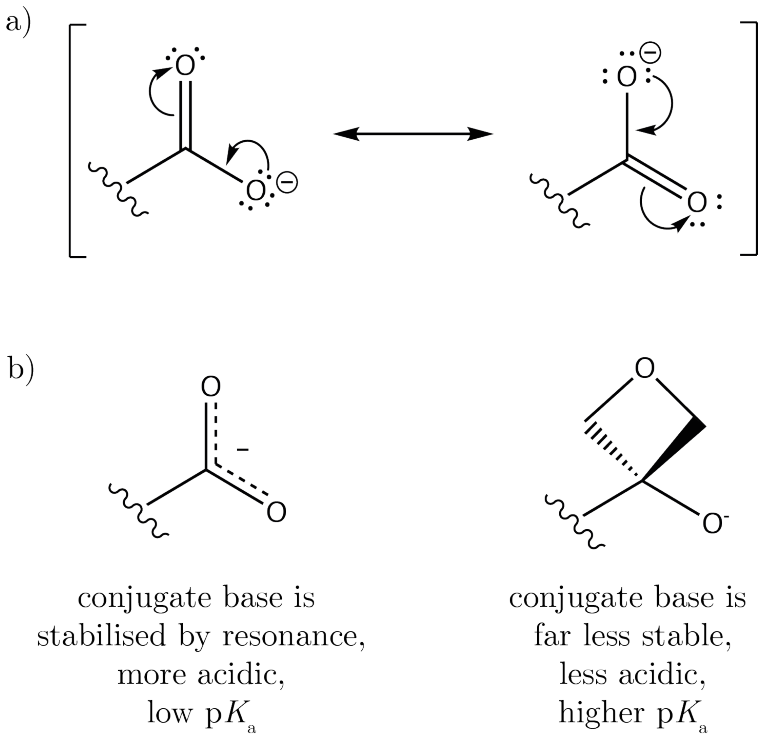
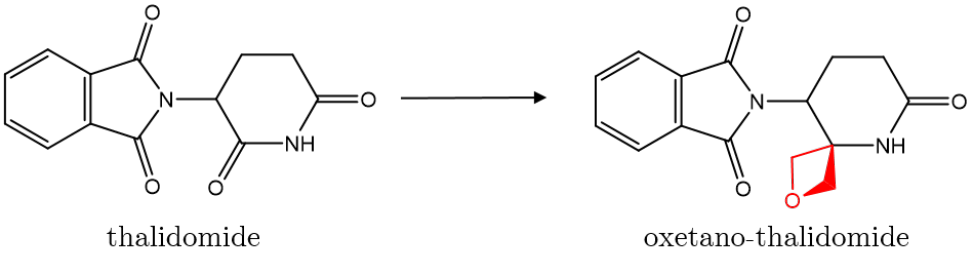


Figure 1.15: Substituting a carbonyl to an oxetane in small molecule drugs results in lower acidity and enhanced permeability, as (a) carboxylate is stabilised by resonance, (b) resulting in a lower pK_a compared to the oxetane-modified molecule.

causing birth defects in up to 12000 children in the 1960s.[61] The two enantiomers of thalidomide readily interconvert *in vivo*, and each enantiomer causes a distinct effect, with (+)-*R* acting as a sedative and (−)-*S* as a teratogen. Substitution of the imide carbonyl with an oxetane ring (Fig. 1.16) produced an analogue which displayed similar physiochemical and *in vitro* properties to the parent thalidomide, but was more stable in human plasma. Importantly, oxetane modification of thalidomide blocks *in vivo* racemisation, suggesting that oxetane modification has the potential to alter teratogenicity (although this work did not include any studies into the biological activity of oxetane-modified thalidomide).[60]



thalidomide oxetano-thalidomide

Figure 1.16: Thalidomide and oxetane-modified analogue. Reproduced from Burkhard et al.[60]

As described, grafting the oxetane motif onto small molecules in place of carbonyl or gem-dimethyl groups can trigger profound changes in metabolic stability, lipophilicity, aqueous solubility and conformational preference without drastically changing hydrogen bonding capabilities or lone pair arrangement (Fig. 1.14).[57,58]

1.4.3 Oxetane-Modified Peptides

In 2014, Shipman[7] and Carreira[62] independently developed a new type of peptide bond isostere, in which the carbonyl of an amide bond is replaced with a 3-aminooxetane unit (Fig. 1.17). Several features of oxetane-modified peptides were expected to make them advantageous as peptidomimetics. Firstly, replacing the amide bond with an oxetane was expected to increase resistance to proteases. Additionally, as the 3-aminooxetane unit has similar hydrogen bonding capabilities and lone pair arrangement to a carbonyl,[53] the oxetane-modified peptidomimetics were expected to support conventional secondary structures. Finally, removing the double bond character of the peptide bond may open up new areas of conformational space that the peptide could sample.

The initial investigation by Shipman and co-workers focussed on the preparation of oxetane-containing tripeptide analogues, in which the central C=O amide bond was substituted for an oxetane ring. The oxetane-modified peptidomimetics were synthesised by conjugate addition of various α-amino esters to a 3-nitromethylene-

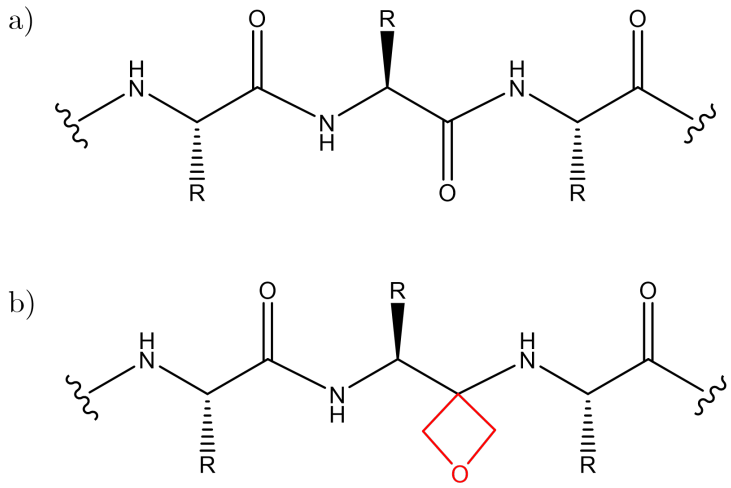


Figure 1.17: (a) Parent and (b) oxetane-modified linear peptide synthesised by the Shipman lab. Reproduced from Beadle et al.[8]

oxetane, reduction of the nitro group followed by coupling with protected amino acids to grow the peptide chain. The structural preferences of these oxetane-modified tripeptides was also studied. In the solid state, oxetane-modified tripeptide Leu-Gly-Ile (LGI) was zwitterionic and displayed antiparallel strand-like arrangements. Both the secondary amine and the oxygen of the 3-aminooxetane unit were involved in the hydrogen bonding network, and further analysis of the crystal structure showed that the nitrogen adjacent to the oxetane is pyramidal with a torsion angle of 60.2°, substantially different from a conventional sp^2-hybridised peptide bond which is planar.[7]

In structures generated from MD simulations, oxetane-modified tripeptide LGI had greater conformational flexibility compared to the unmodified parent, and the C- and N-termini of the oxetane-modified peptides were closer together.[7] This led to the hypothesis that oxetane could increase efficiency of cyclisation by inducing a turn in the peptide backbone. The Shipman group have since produced a variety of linear and cyclic oxetane-modified peptides using solid-phase peptide synthesis methods.[8]

The synthetic approach to oxetane-modified peptides, developed independently by Shipman[7] and Carreira,[62] is highly efficient for the preparation of peptidomimetics in which the oxetane residue is based on glycine (G_{ox}). Subsequent work by Shipman and co-workers involved the preparation of oxetane-modified alanine (A_{ox}) residues.[9,10] To date, it is not possible to prepare oxetane-modified structures in which the modification is based on other amino acid residues. This could in theory be achieved using work published by Carreira,[63] but the approach is lengthy and challenging.[41]

By using the oxetane modification, it may be possible to much more read-

ily chemically synthesise cyclic peptides, which have potential as useful drug molecules, thanks to their improved physicochemical properties compared to their linear counterparts. Therefore, it is important to investigate the effect of oxetane modification on small peptides, to better understand how oxetane could increase macrocyclisation efficiency, and on the structure of relevant biological molecules, such as α-helical structures.

1.5 Azetidine Modification

The main limitation of oxetane modification is the poor stability of the oxetane ring under acidic conditions.[53] This can be overcome to a degree by altering the protecting groups used and only employing mildly acidic conditions.[64] However, preparing oxetane-modified peptides containing residues with protecting groups which require harshly acidic conditions, such as arginine (particularly important for antimicrobial activity) is still a significant challenge.[41] Instead, alternative heterocycles such as azetidine could yield similar advantages in the preparation of cyclic peptides. Like oxetane, azetidine could help bring the termini of a peptide closer together, and 3-aminoazetidine-3-carboxylic acid (Azt) residues containing an azetidine ring covalently bonded to the Cα have been reported to induce a type II β-turn.[65]

The azetidine heterocycle is similar to oxetane, but contains an NH in place of the oxygen atom (Fig. 1.18). Azetidines have been of interest in medicinal chemistry since the 1970s,[66] they are able to enhance metabolic stability and pharmacokinetic properties[67–69] and are found in several drugs (e.g. Baricitinib, Cobimetinib)[70,71] and biologically active natural products (e.g. Gelsemoxonine).[72] They have attracted new attention due to more recent work involving functionalisation of amines in strained rings.[73,74] By selective functionalisation of the azetidine nitrogen, it may be possible to prepare functional derivatives with a variety of applications.[75–77] An additional benefit of azetidine-modified peptides is that the azetidine ring is stable under acidic conditions, unlike oxetane, which is prone to ring-opening under acidic conditions.[53] This is because the C-N bond is less polarised and the partial positive charge on the carbon is smaller as nitrogen is less electronegative than oxygen, making it less susceptible to attack by water (Fig. 1.19). The instability of oxetanes in acidic conditions limits the scope of amino acids that oxetane-modified peptides can contain, as all of the amino acid sidechain protecting groups used are removable using acids,[64] whereas this is not a problem in the preparation of azetidine-modified peptidomimetics.

Both oxetane- and azetidine-modified peptides have exciting implications as

potential new therapeutics.

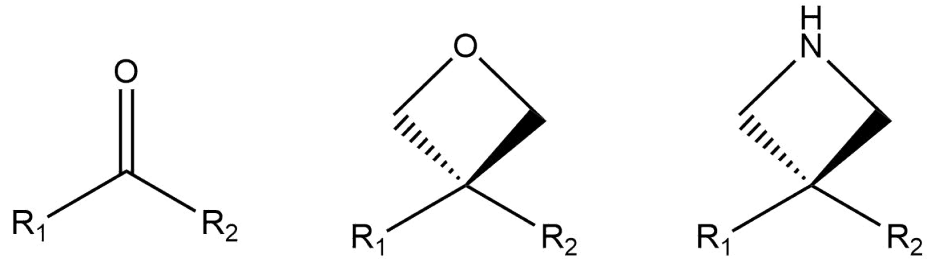


Figure 1.18: Comparison of carbonyl group, oxetane ring and azetidine ring.

1.6 Antimicrobial Peptides

Antimicrobial peptides (AMPs) are increasingly promising treatment strategies for bacterial infections. Antibiotic resistance is a global crisis that threatens life as we know it. Pathogenic resistance to antibiotics continues to spread due to inappropriate usage in healthcare and agriculture, and pathogens resistant to one antibiotic must be treated with another, resulting in the emergence of multi-drug resistant bacteria. Of particular concern is the prevalence of multi-drug resistance in Gram-negative species such as *Eschericia coli* and *Acinetobacter baumannii*, where pan-drug resistance (i.e. resistance to all classes of antibiotics) has been observed.[78] Drug-resistant diseases already cause over 700,000 deaths a year globally, and unless action is taken this figure could increase to 10 million deaths per year globally by 2050,[79] with increasing concerns about a 'post-antibiotic' era with no antimicrobial treatment options.[80]

AMPs are part of the innate immunity of most living organisms (from bacteria to vertebrates),[80] and are generally small, often with an overall net positive charge. The most well-known examples of AMPs include polymyxin B, a cationic peptide produced by *Bacillus polymyxa*, and vancomycin, a non-cationic glycopeptide produced by *Amycolatopsis orientalis*. Both polymyxin B and vancomycin are FDA-approved antibiotics.[81]

AMPs can exhibit antibacterial, antifungal, antitumour and antiviral activity, and, importantly, they have potential to overcome bacterial resistance.[82] AMPs often have broad spectrum activity against Gram-positive and Gram-negative bacteria, and many of them rapidly act on the cytoplasmic membrane (rather than on specific proteins, like most antibiotics), resulting in a decreased rate of resistance being developed.[83,84]

Figure 1.19: Comparison of acid stability between (a) oxetane, and (b) azetidine. Oxygen is highly electronegative, resulting in a partial positive charge on the carbon, making it susceptible to attack by water. As nitrogen is less electronegative, the C-N bond is less polarised than a C-O (with a smaller partial positive charge on the carbon), making it less susceptible.

There are several modes of action that have been proposed for AMPs, including those which involve selective interaction of the AMP with the cell membrane of the target organism, which is highly dependent on the structure of the peptide,[82] although AMPs may also target intracellular processes such as DNA and protein synthesis.[85]

The most widely accepted membrane-active mechanistic models are the barrel stave model, the carpet model, and the toroidal pore model (Fig. 1.20). The barrel stave model describes the formation of transmembrane pores by amphipathic α-helices, where the hydrophobic surfaces of the peptides are interacting with the lipid core of the membrane.[82] The carpet model involves the peptides lining up parallel to the membrane surface and forming a peptide carpet, which is followed by detergent-like action resulting in pore formation.[86] The toroidal pore model is similar to the barrel stave model in that it involves formation of transmembrane pores by peptides, but the pores are lined with peptide and lipid.[85]

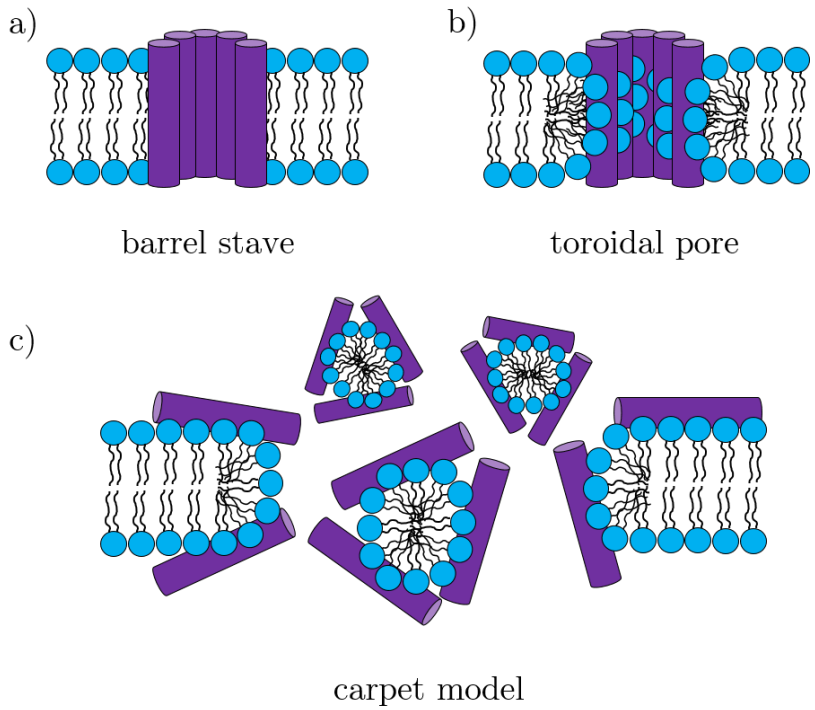


Figure 1.20: Classical models of membrane disruption by antimicrobial peptides, in which peptides reach a threshold concentration sufficient for them to insert across the membrane to (a) form peptide-lined pores in the barrel stave model, (b) form peptide-and-lipid-lined pores in the toroidal pore model, or (c) solubilise the membrane into micellar structures in the carpet model. Adapted from Nguyen et al.[85]

There are several factors that are thought to be important for the antimicrobial activity of AMPs, such as net charge (most are highly cationic), hydrophobicity, amphipathicity and secondary structure. Additionally, specific amino acids also

result in increased antimicrobial activity.[87] In particular, arginine and tryptophan residues are found in high proportions in many AMPs. These residues are thought to increase antimicrobial activity due to their positive charge (Arg) and hydrophobic bulk (Trp).[86]

Cationic residues (Arg and Lys) are thought to mediate the initial electrostatic attraction to the negatively charged bacterial cell membranes.[87] Arginine appears to have an advantage over lysine, in that when in a stacked cation-π arrangement with tryptophan, the Arg sidechain is still able to form almost as many hydrogen bonds with surrounding water molecules as when it is not involved in any cation-π interactions.[88] Lysine, however, cannot form hydrogen bonds when forming cation-π interactions with an aromatic residue. These cation-π interactions effectively shield the Arg from the hydrophobic lipid bilayer, making entry more energetically favourable.[86] Additionally, Trp residues have a preference for the interfacial region of lipid bilayers, anchoring the peptide to the bilayer surface and allowing for prolonged association with the membrane.[87]

As previously discussed, cyclic peptides often have enhanced biological activity compared to their linear counterparts, and there have been several studies which suggest that antimicrobial activity can be enhanced by cyclisation.[89–91] Compared to linear peptides, cyclic peptides demonstrate enhanced stability, low toxicity and enhanced binding properties, making cyclic AMPs attractive potential therapeutics and a growing area of interest.[89]

1.7 Aims and Objectives

The aim of this study was to use a combination of biophysical, microbiological and computational techniques to understand the impact of oxetane and azetidine modification on small peptides and peptides of helical secondary structure. This has important implications in the development of potential therapeutics, as oxetane and azetidine modification have previously been used in medicinal chemistry,[53,66] and have the potential to aid in cyclisation of small peptides,[7] which has been shown to improve pharmacokinetic properties.[16] This new understanding was then applied to a biologically active peptide, specifically an AMP, to measure the effect of modification on biological activity. In order to achieve this aim, the study was divided into three smaller objectives:

- Characterise the impact of oxetane modification on small peptides (cyclic and linear) using a combination of biophysical and computational techniques.
- Characterise the impact of oxetane modification on longer peptides with α-helical content (an important motif in biology), using a predominantly

computational approach.

- Assess changes in antimicrobial activity following cyclisation and/or azetidine modification for a model sequence with known biological activity.

The first and second objectives described are extending upon previous work by Powell et al.[7] and Beadle[9] respectively. Previous work has involved the preparation and simulation of linear oxetane-modified tripeptides, so we aimed to study longer (pentapeptide) sequences, both linear and cyclic, using simulations and complementary biophysical techniques. The effect of oxetane modification on helical peptides had previously been studied using circular dichroism, but the mechanism by which oxetane reduces helical content was not understood, therefore we aimed to characterise this using a computational approach. The third objective was completely novel—we wanted to characterise how azetidine modification of peptides alters biological activity, as this had never been done before and has important implications in the development of potential therapeutics.

1.8 Chapter Organisation

In Chapter 2, simulation theory is outlined. Methodologies (both experimental and computational) are reported in Chapter 3. The results are split into three chapters. The impact of oxetane modification on small peptides (linear and cyclic), as described using a combination of biophysical techniques and molecular dynamics (MD) simulations, is presented in Chapter 4, and this work is in part published in Roesner et al.[92] Chapter 5 focusses on using steered MD simulations to simulate the unwinding of helical peptides, to assess differences between unmodified and oxetane-modified residues, and to compare to circular dichroism data. This work is published in Jayawant et al.[93] The effect of azetidine modification on antimicrobial activity is described in Chapter 6. Each of the results chapters contain an introduction section, a results and discussion section, and finish with a conclusion. Finally, Chapter 7 discusses overall conclusions and future directions for the work as a whole.

Chapter 2

Theory

2.1 Statistical Mechanics

Many experimental techniques aim to measure thermodynamic macroscopic bulk properties, time averaged over a large number of particles. However, as bulk measurements, these techniques cannot convey the full picture of a system. In order to elucidate the components that make up the observed bulk property, it is necessary to 'zoom in' to the atomic scale, and simulations (sometimes referred to as the 'computational microscope'[94]) are an increasingly popular tool for this purpose. Since classical computational simulations calculate the positions and velocities of atoms, the data cannot be compared directly to the averaged bulk properties obtained using experimental techniques.[95] Statistical mechanics involves the use of probability theory to link the microscopic information from simulations to the macroscopic experimental data.[96] A macroscopic property, A, will depend on the position and momenta of the N particles which make up the system, and the value of A changes over time due to the way the particles interact in the system. Macroscopic systems consist of huge numbers of particles (in the order of 10^{23}, Avogadro's number), and it is not feasible to calculate the way every particle moves in a time-dependent manner, since today's simulations are typically only in the order of 10^6 particles. Instead statistical mechanics relies on the concept of microstates and ensembles for a system.

A microstate is a microscopic configuration of the N atoms in a system. Each point in the phase space of a system is a microstate, and can be expressed by a set of values $6N = (r_{x1},\ r_{y1},\ r_{z1},\ ...,\ r_{xN},\ r_{yN},\ r_{zN},\ p_{x1},\ p_{y1},\ p_{z1},\ ...,\ p_{xN},\ p_{yN},\ p_{zN})$, where r is the position and p is the momentum. Thus the phase space of a thermodynamic system is $6N$-dimensional, with three position co-ordinates and three components of momenta for each atom. An ensemble is a collection of all possible microstates with identical macroscopic properties. An example of this

could be an identical arrangement of particles in different regions of a box. Their atomic co-ordinates and configurations are different, but the total energy of the two microstates is identical. In statistical mechanics, a single system evolving in time is replaced by a large number of replications of the system that are considered simultaneously.[96] Therefore, the time average of single system can be replaced by the ensemble average, $\langle A \rangle_{\text{ensemble}}$, for multiple microstates at a given time point:

$$\langle A \rangle_{\text{ensemble}} = \iint dp^N dr^N A(p^N, r^N) \rho(p^N, r^N) \tag{2.1}$$

where $A(p^N, r^N)$ is the macroscopic property and $\rho(p^N, r^N)$ is the probability of finding the system in which the particles are in a microstate with particle momenta p^N and positions r^N.

Ergodicity is used to describe a system in which the time averaged behaviour across all microstates equals the ensemble average:

$$\langle A \rangle_{\text{ensemble}} = \langle A \rangle_{\text{time}} \tag{2.2}$$

This is important in MD, because a time average is computed and used to calculate experimentally relevant macroscopic properties.

2.2 Interaction Potentials

Quantum mechanics calculations can provide information about the behaviour of a system on an atomic and subatomic level, but are so computationally expensive that it is only possible to study a small number of particles. As biologically relevant systems are much larger (on the scale of 10^{23} atoms), it is necessary to use alternative means to model these systems. Molecular mechanics, or force field theory, involves stripping away detail by ignoring the electronic motions of atoms and allows for the simulation of many-body systems. Although information about quantum effects is lost, molecular mechanics provides an excellent approximation for a wide range of materials.[95]

In force field theory, the interaction between particles or atoms is described using a set of parameters which relate the chemical structure to the total energy of the system. A force field is the set of interaction potentials and parameters that describe the forces acting on each type of atom in a system due to its interactions with the other atoms in the system. It is generally broken down into non-bonded and bonded potentials. Bonded interactions include bond stretching, angle bending and torsion angle rotation, and together describe the energy penalties associated with the movement of bonds and angles away from equilibrium. Non-bonded

interactions include van der Waals forces and electrostatic interactions, and these describe how atoms that are not covalently bonded together interact through space as a function of distance. The potential energy, V, of a system can be described as a sum of these interaction components as a function of the positions of the particles, r, which is represented by the force field's functional form:

$$\begin{aligned} V(r) &= \sum_{bonded} K_{bonded}(r) + \sum_{non-bonded} K_{non-bonded}(r) \\ &= \underbrace{\sum_{bonds} K_{bonds}(r) + \sum_{angles} K_{angles}(r) + \sum_{dihedrals} K_{dihedrals}(r)}_{\text{Bonded interactions}} \\ &+ \underbrace{\sum_{vdW} K_{vdW}(r) + \sum_{electrostatics} K_{electrostatics}(r)}_{\text{Non-bonded interactions}} \end{aligned} \tag{2.3}$$

where parameters are defined in the following section, and are represented visually in Fig. 2.1. The way these parameters will have been determined depends on the force field, but are generally based on *ab initio* calculations or experimental observations. More specialised force fields may contain additional terms, but the most frequently used force fields for biological molecules, including CHARMM27,[97] take this form.

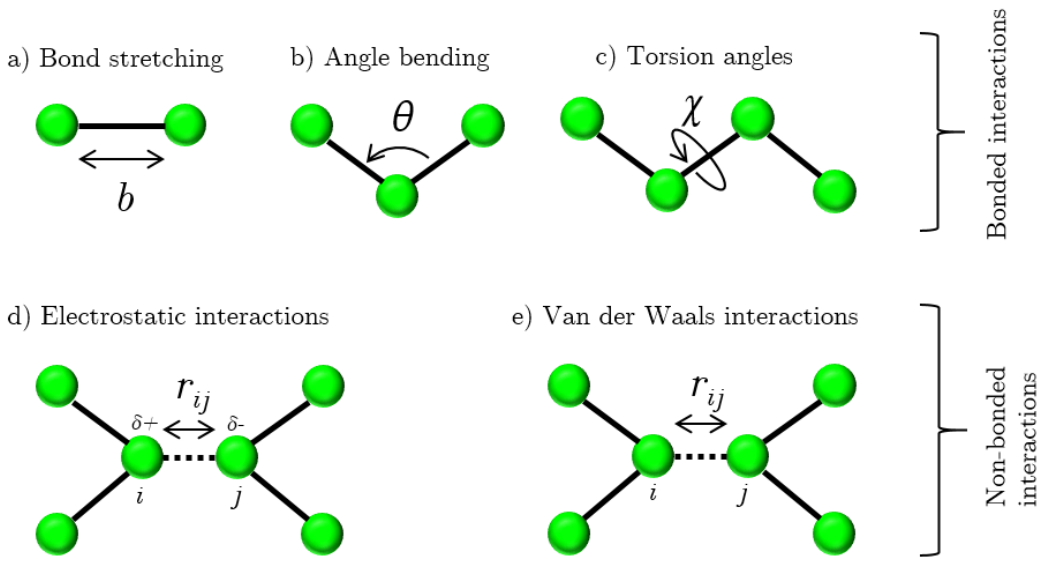


Figure 2.1: Representation of the most common interaction components that comprise a basic force field. These components are (a) bond stretching, (b) angle bending, (c) torsion angles, (d) electrostatic interactions, and (e) Van der Waals interactions.

2.2.1 CHARMM Force Field

The force field predominantly used in this work is the Chemistry at Harvard Macromolecular Mechanics (CHARMM) force field. CHARMM is a popular force field for the simulation of biomolecules. The CHARMM potential energy function[98] is described below:

$$\begin{aligned} V(r) = & \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\chi[1+\cos(n\chi-\delta)] \\ & + \sum_{Urey-Bradley} K_{UB}(S-S_0)^2 + \sum_{impropers} K_\phi(\phi-\phi_0)^2 \\ & + \sum_{\substack{non-bonded \\ atom\ pairs}} \left(\varepsilon \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_D r_{ij}} \right) \end{aligned} \tag{2.4}$$

where V is the potential energy, r is the ensemble of co-ordinates of atoms in the system. K_b, K_θ, K_{UB}, K_χ and K_ϕ represent the bond, valence angle, Urey-Bradley (U-B), dihedral angle, and improper dihedral angle force constants, respectively. b, θ, S, χ, and ϕ are the bond length, bond angle, U-B 1,3 distance, dihedral torsion angle, and improper dihedral angle, respectively. The subscript zero indicates the equilibrium values for each term. n represents the multiplicity and δ is the phase of the dihedral angle (see section 2.2.4). Collectively, these terms are referred to as the internal terms in the potential energy function. ε is the Lennard-Jones (L-J) well-depth and $R_{\min}$ is the distance at the L-J minimum interaction energy (see section 2.2.6). q_i represents the partial atomic charge and ε_D is the dielectric constant (see section 2.2.5). r_{ij} is the distance between atoms i and j, respectively. Given the position of the atoms, r, all the distances and angles required to evaluate $V(r)$ can be readily determined.[97]

There are a couple of terms in the CHARMM functional form which are absent from the standard molecular mechanics force field described previously. The U-B term is a harmonic term describing the distance between atoms 1 and 3 of some of the angle terms. The U-B term was introduced to optimise the fit to vibrational spectra.[99] Additionally, there is a separate term for improper torsion angles. While a dihedral angle potential depends on four consecutive bonded atoms, improper torsion depends on three atoms located around a fourth atom, and describes out-of-plane bending motions (Fig. 2.2).

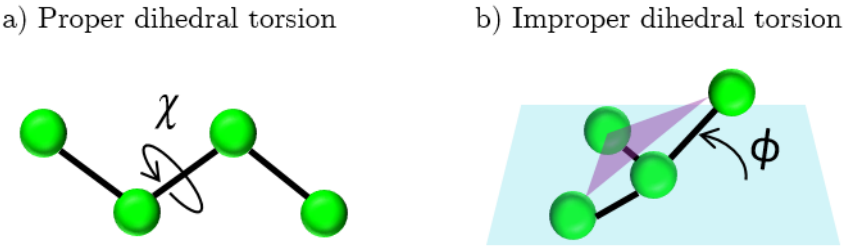


Figure 2.2: Representation of dihedral torsion angles. Proper dihedral angles (a) depend on four consecutively bonded atoms, and improper dihedral angles (b) involve three atoms surrounding around a fourth.

2.2.2 Bond Stretching

The potential energy curve for a typical bond has a form which can be described by a Morse potential:

$$V_{bond}(b) = D_e\{1 - \exp[-a(b - b_0)]\}^2 \tag{2.5}$$

D_e is the depth of potential energy minimum and $a = \omega\sqrt{\mu/2D_e}$, where μ is the reduced mass and ω is the frequency of bond vibration. ω is related to the stretching constant of the bond, k, by $\omega = \sqrt{k/\mu}$. b_0 is the reference value of the bond (the value that the bond adopts when all other terms in the force field are set to zero). This Morse potential is not usually used in molecular mechanics force fields, for several reasons. It is not very amenable for efficient computation, in part because it requires three parameters to be specified for each bond. Additionally, the Morse curve describes a wide range of behaviour, from strong equilibrium to dissociation. Since it is rare for bonds in molecular mechanics calculations to deviate significantly from their equilibrium values, much of the information described by a Morse curve is unnecessary, and thus simpler approximations can often be used instead.[96] The simplest commonly used approximation of a bond is described by Hooke's law, which is a harmonic potential:

$$V_{bond}(b) = \frac{k_b}{2}(b - b_0)^2 \tag{2.6}$$

where $V_{bond}(b)$ is the bond energy as a function of the bond length, b, k_b is the bond force constant, and b_0 is the reference bond length. The CHARMM27 force field uses K_b rather than $\frac{k_b}{2}$, this is because the K_b given in the parameter files is half the value of a traditional spring constant.

In proximity to the energy minimum, the Morse potential and harmonic potential are in close agreement (Fig. 2.3). The CHARMM27 force field (which is

used throughout this work) uses simple harmonic potentials to represent bond lengths. Alternatively, bond lengths can be constrained using various constraint algorithms, which are discussed in further detail in a subsequent section.

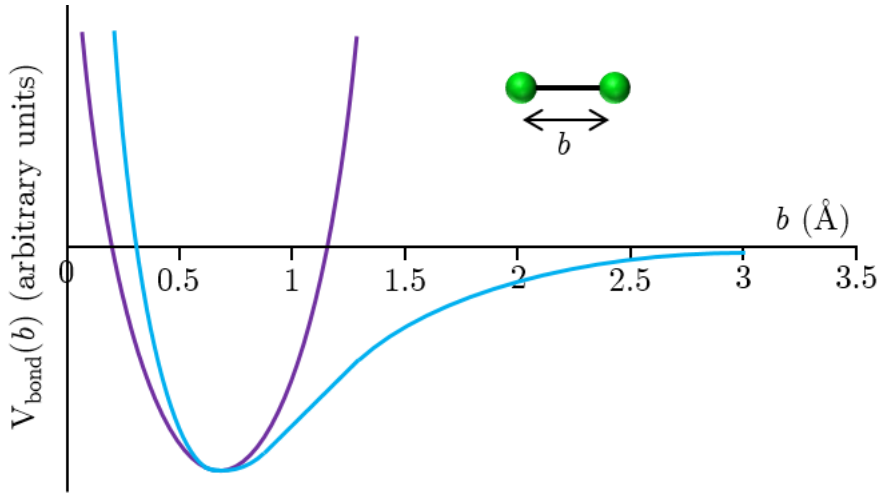


Figure 2.3: Representation of bond energy as a function of length, described by the harmonic potential (purple) and the Morse potential (blue).

2.2.3 Angle Bending

A harmonic potential is also used frequently to describe the deviation of angles from their reference value:

$$V_{angle}(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{2.7}$$

where $V_{angle}(\theta)$ is the energy as a function of the angle, θ, formed between the three covalently bonded particles, K_θ is the angular force constant, and θ_0 is the equilibrium angle. As with bond stretching, CHARMM27 force field uses K_θ rather than $\frac{k_\theta}{2}$, the halving is incorporated already into the force field.

2.2.4 Torsion Angle Rotation

The bond stretching and angle bending terms in the force field require substantial energies in order to deviate significantly from their reference values. Most of the variation in relative energies is due to changes in torsion angles and non-bonded contributions.[96] In biomolecules, major changes in conformation are often due to rotations about bonds (consider the classic boat and chair conformations of glucose). In order to simulate these molecules it is vital for the force field to accurately represent the energy profile of such changes.

Torsion potentials can be represented as a cosine series expansion:

$$V_{dihedral}(\chi) = \sum_n \frac{V_n}{2}[1 + \cos(n\chi - \delta)] \tag{2.8}$$

where χ is the torsion angle, V_n is the barrier height (provides a qualitative indication of the relative barriers to rotation), n is the multiplicity (gives the number of minimum points in the function as the bond is rotated through 360°), and δ is the phase factor (determines where the torsion angle passes through its minimum value).

As previously mentioned, the CHARMM27 force field also contains a term to describe improper torsions. In certain molecular components, for example the groups forming an aromatic ring, the atoms are always approximately planar. Improper dihedral angle terms in a force field use an out-of-plane bending component to enforce planarity, thereby imposing the correct geometry on these specific groups. A torsion potential can be used to maintain the improper torsion angle at 0° or 180°:

$$V_{improper}(\phi) = k(1 - \cos 2\phi) \tag{2.9}$$

Alternatively, a harmonic form can be used, as in CHARMM27:

$$V_{improper}(\phi) = \frac{k_\phi}{2}(\phi - \phi_0)^2 \tag{2.10}$$

2.2.5 Electrostatic Interactions

Electrostatic interactions are through-space interactions that occur between non-bonded, charged particles. The electrostatic interaction between two non-bonded molecules is calculated using Coulomb's law:

$$V_{electrostatics}(r_{ij}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r r_{ij}} \tag{2.11}$$

where $V_{electrostatics}(r_{ij})$ is the Coulombic interaction energy as a function of distance r_{ij} between particles i and j, ε_0 is the vacuum permittivity, and q is the charge.

The Coulomb interaction energy decays slowly, and covers a long range that is often larger than the size of the system (Fig. 2.4). Therefore, long-range electrostatic interactions make an important contribution to the energy and must be considered, so it is not possible to introduce a simple cut-off to reduce computational costs as is done for L-J interactions (see section 2.2.6). Instead, more sophisticated methods are used.

One such method is the Particle Mesh Ewald (PME) summation,[100] an implementation of the Ewald summation method.[101] In a periodic system (see section

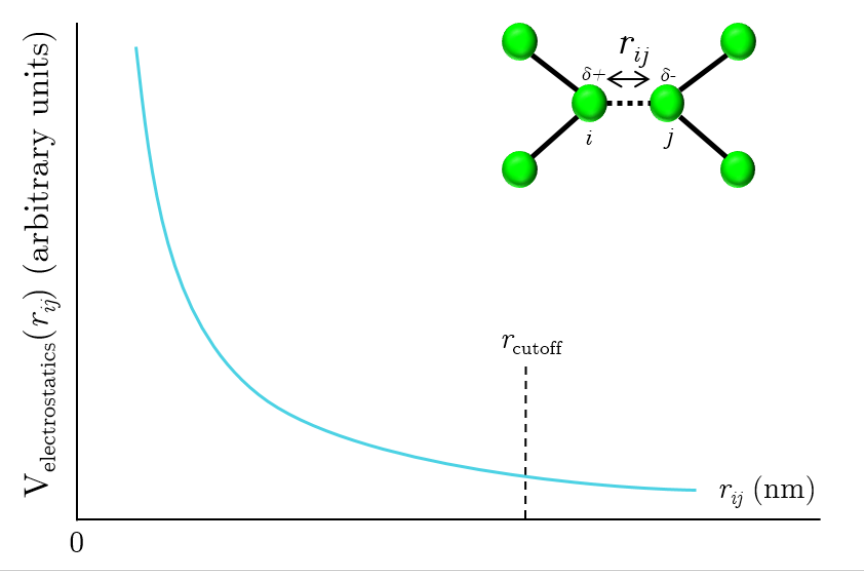


Figure 2.4: Representation of the Coulomb interaction between two atoms i and j. The interaction energy decays slowly, so truncating the potential at a cut-off value (r_{cutoff}) has a significant impact on long-range interactions, highlighting the need for the Ewald summation method.

2.4.4), particles interact with every other particle within the simulation box, as well as their periodic images, in an infinite lattice. The total electrostatic energy can be derived from the sum of these contributions, but it is very slow to converge upon a solution. The Ewald summation method breaks down the summed electrostatic interactions into short-range (real space particle-particle) and long-range (Fourier space) components. The PME method is much faster compared to standard Ewald summation, as it uses Fast Fourier Transforms (FFT), in which the charges are assigned to grid points and an FFT algorithm is implemented to give the reciprocal energy. The total long-range interaction energy is then calculated using an inverse Fourier transformation and summation over the grid points. This method is much more efficient for use in large systems, as the computational cost of the PME algorithm scales as $N \log N$ (compared to N^2 for Ewald summation).

2.2.6 Van der Waals Interactions

Electrostatic interactions do not make up the entirety of the non-bonded interactions occurring between particles, as forces also occur between electrically neutral atoms, due to fluctuations in the positions of electrons. These fluctuations induce attractive forces due to dipole-dipole interactions at long-range separations, and repulsive forces due to overlap of electron distributions at short-ranges. Collectively, these intermolecular interactions are called van der Waals forces, and these

can be approximated by an L-J 12-6 potential:

$$V_{vdW}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{2.12}$$

where $V_{vdW}(r_{ij})$ is the energy as a function of the intermolecular separation r_{ij} between two particles i and j, ε and σ describe the depth of the potential well and the collision diameter (the distance between pairwise particles i and j at which the potential energy is zero, Fig. 2.5) respectively.

This can be related back to the form shown in Eqn. 2.4, where:

$$R_{\min} = 2^{1/6}\sigma \tag{2.13}$$

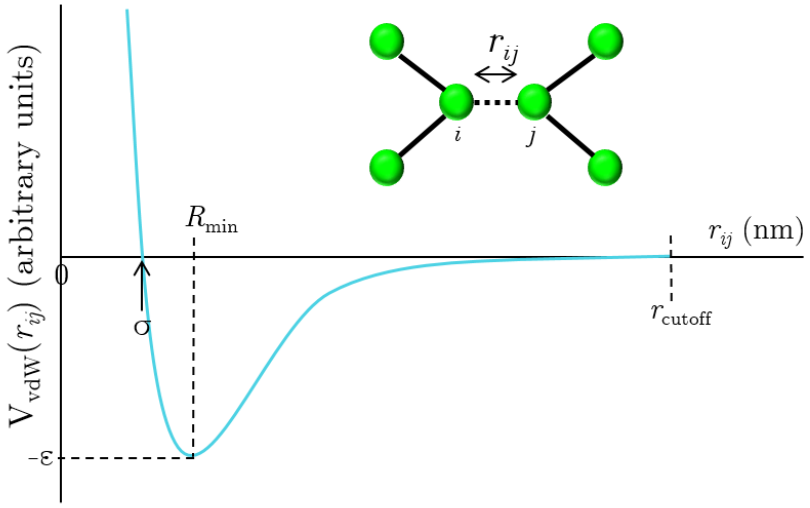


Figure 2.5: Representation of an L-J potential between two atoms i and j. At the minimum of the energy well, $R_{\min}$, the potential energy equals ε. Since the plot tails off rapidly, implementing a cut-off value (r_{cutoff}) has a negligible effect on interactions.

2.3 Energy Minimisation

Unless the system is very small and simple, the potential energy is a complicated and multi-dimensional function of the co-ordinates. For a system with N particles the energy is a function of $3N$ Cartesian co-ordinates, and thus is impossible to visualise.[96] Locations of minimum energy exist on this multi-dimensional surface, which are local energy minima, and the lowest point overall is the global energy minimum.

In this study, peptide models were initially built in either a linear, cyclic or he-

lical fashion by arranging the atoms relative to each other based on the chemical structures of the individual amino acids, and were then solvated. It is necessary to minimise the potential energy between the atoms to ensure the starting geometry of the model is optimal. Steric clashes between atoms can be introduced during the initial creation of a system, and the proximity between particles can result in high intermolecular forces that cause the time step integrator to fail (the system 'explodes'). To help prevent steric clashes, the potential energy between atoms can be minimised through the use of energy minimisation algorithms, which move the system 'downhill' on the energy surface to the nearest local energy minimum (Fig. 2.6). Three of the most commonly used potential energy minimisation algorithms are steepest descent,[102] conjugate gradients,[103] and limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) quasi-Newtonian minimizer.[104–106] In this work, all starting systems were energy minimised using the steepest descent algorithm.

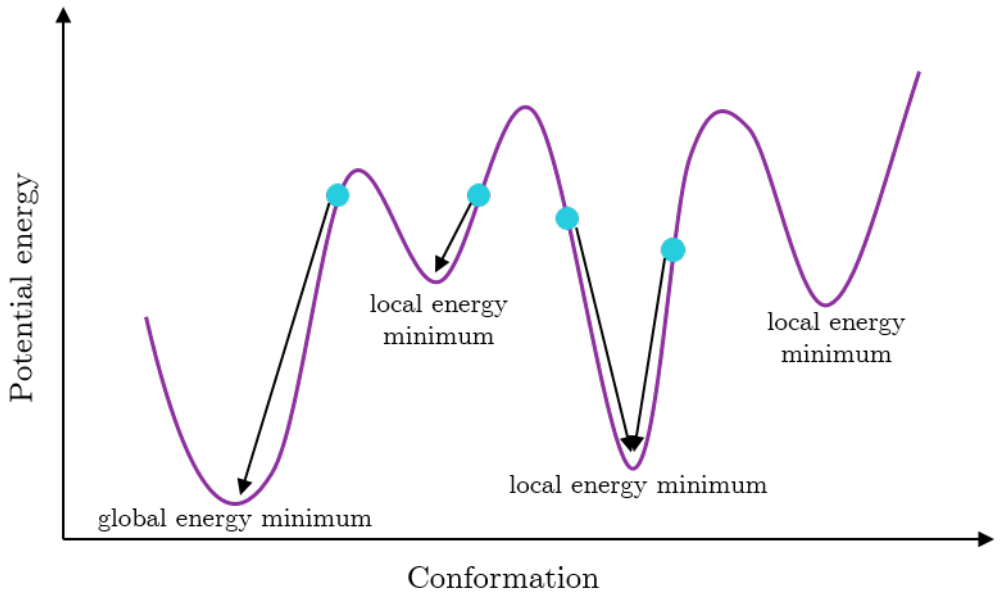


Figure 2.6: Representation of a one-dimensional potential energy surface. Energy minimisation algorithms move the system downhill (represented with arrows) from the starting co-ordinates (blue circles) to the nearest local energy minimum. This may not be the global energy minimum.

The steepest descent algorithm is a first order optimisation algorithm which iteratively adjusts the co-ordinates of the atoms to move the system down the steepest local energy gradient. Initially, a maximum displacement value, h_0 is chosen (e.g. 0.01 nm). First the forces, F, and potential energy are calculated.

New positions are calculated using the following:

$$r_{n+1} = r_n + \frac{F_n}{\max(|F_n|)} h_n \tag{2.14}$$

where r_n and r_{n+1} represent the vector of all $3N$ co-ordinates at the n^{th} and $n+1^{\text{th}}$ iteration steps, F_n is the force, $\max(|F_n|)$ is the largest scalar force on any atom, and h_n is the maximum displacement. The forces and energy are calculated before (V_n) and after (V_{n+1}) each iteration. If $V_{n+1} < V_n$, the new co-ordinates are accepted and the displacement value is increased ($h_{n+1} = 1.2h_n$). If $V_{n+1} > V_n$, the new positions are rejected and the displacement value is reduced ($h_n = 0.2h_n$). The algorithm stops either after a pre-defined number of steps or when $\max(|F_n|)$ falls below a specified value ε. The value for ε can be estimated from the root mean square force f a harmonic oscillator exhibits at temperature T:

$$f = 2\pi v\sqrt{2mkT} \tag{2.15}$$

where v is the oscillator frequency, m is the reduced mass, and k_B is Boltzmann's constant.

Stopping the algorithm once $\max(|F_n|)$ falls below ε ensures that the system is within close vicinity of a local minimum. The steepest descent method was used to minimise the potential energy of all systems prior to running the simulations.

2.4 Molecular Dynamics

Molecular dynamics (MD) is a technique for simulating the motions of particles in a system, and simulations have become a standard tool for probing the natural structure and dynamics of biomolecules.[107]

In order to perform a MD simulation, it is essential to have a knowledge of the interaction potential for the particles, and of the equations of motion governing the dynamics of the particles.[108] For biomolecular systems, classical Newtonian equations of motion are suitable. These three physical laws describe the relationship between a body and the forces acting upon it, and its motion in response.[96]

These three laws can be summarised as follows:

1. An object moves in a straight line at constant velocity unless a force acts upon it.

2. The force on an object is equal to the rate of change of momentum, which

can be written as the following equation:

$$F = ma \tag{2.16}$$

where F is the force exerted, m is the mass of an object or particle, and a is the acceleration.

3. For every action there is an equal and opposite reaction.

In MD, successive configurations of the system are generated by integrating Newton's laws of motion, resulting in a trajectory, which specifies how the positions and velocities of particles in the system vary with time (Fig. 2.7).

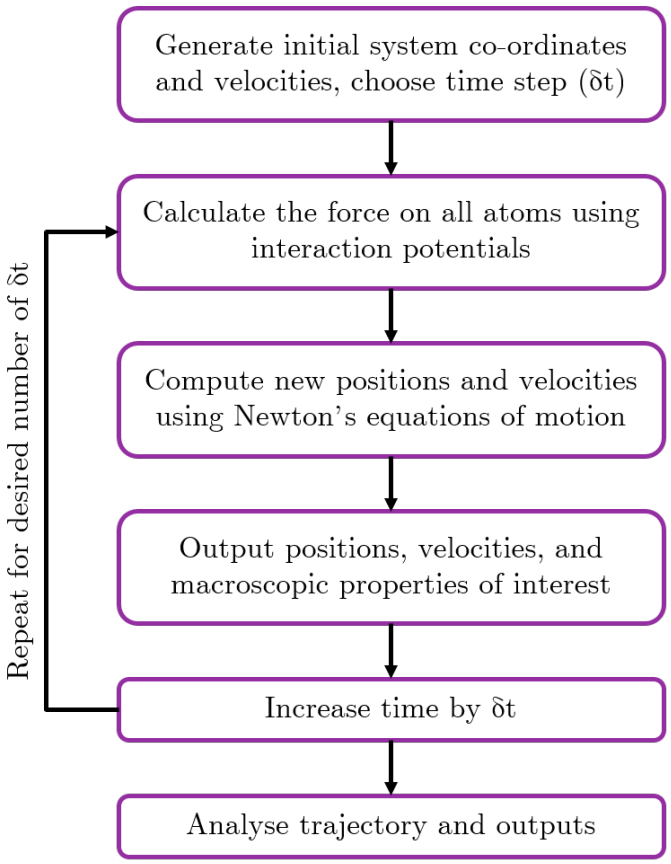


Figure 2.7: The basic MD algorithm for a standard MD simulation.

The trajectory produced by a MD simulation is obtained by solving the differential equations described by Newton's second law:

$$\frac{d^2}{dt^2} r_i(t) = \frac{d}{dt} v_i(t) = a_i(t) = \frac{F_i(t)}{m_i} \tag{2.17}$$

where $r_i(t)$ is the position, m_i is the mass, $a_i(t)$ is the acceleration, $v_i(t)$ is the velocity and $F_i(t)$ is the force on particle i at time t.

It is necessary to use a numerical integration algorithm, because the force on each atom changes over the trajectory and it is coupled to the force acting on every other atom (i.e. it is a many body problem), and cannot be solved analytically.

2.4.1 Leap-Frog Algorithm

The leap-frog algorithm[109] is a time integration method used to integrate the equations of motion in an MD simulation. It is a modified version of the Verlet algorithm,[110] which uses the positions and accelerations at time t, and the positions from the previous step, $r(t - \delta t)$, to calculate new positions at time $t + \delta t$, $r(t + \delta t)$. The positions at some small time step δt from the current time t can be approximated by a Taylor expansion:

$$r_i(t + \delta t) = r_i(t) + \delta t v_i(t) + \frac{1}{2}\delta t^2 a_i(t) + ... \tag{2.18}$$

$$r_i(t - \delta t) = r_i(t) - \delta t v_i(t) + \frac{1}{2}\delta t^2 a_i(t) - ... \tag{2.19}$$

where $r_i(t)$ is the position, m_i is the mass and $a_i(t)$ is the acceleration of particle i at time t. Substituting these two equations together gives:

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + \delta t^2 a_i(t) \tag{2.20}$$

The velocity of particle i at the half-step, $t + \frac{1}{2}\delta t$, can be estimated by:

$$v_i(t + \frac{1}{2}\delta t) = \frac{r_i(t + \delta t) - r_i(t)}{\delta t} \tag{2.21}$$

The leap-frog algorithm uses the following relationships:

$$r_i(t + \delta t) = r_i(t) + \delta t v_i(t + \frac{1}{2}\delta t) \tag{2.22}$$

$$v_i(t + \frac{1}{2}\delta t) = v_i(t - \frac{1}{2}\delta t) + \delta t a_i(t) \tag{2.23}$$

$$v_i(t) = \frac{1}{2}[v_i(t + \frac{1}{2}\delta t) + v_i(t - \frac{1}{2}\delta t)] \tag{2.24}$$

Based on these equations, the particle velocity at time $t + \frac{1}{2}\delta t$ can be calculated based on the velocity at time $t - \frac{1}{2}\delta t$ and the acceleration at time t. The positions then 'leap-frog' over the velocities to give their new values at $t + \delta t$, hence the name 'leap-frog algorithm' (Fig. 2.8).

Choosing an appropriate time step is extremely important in MD simulations. If the time step is too small, the trajectory will not sample enough phase space, and if the time step is too large then instabilities can arise in the integration algorithm

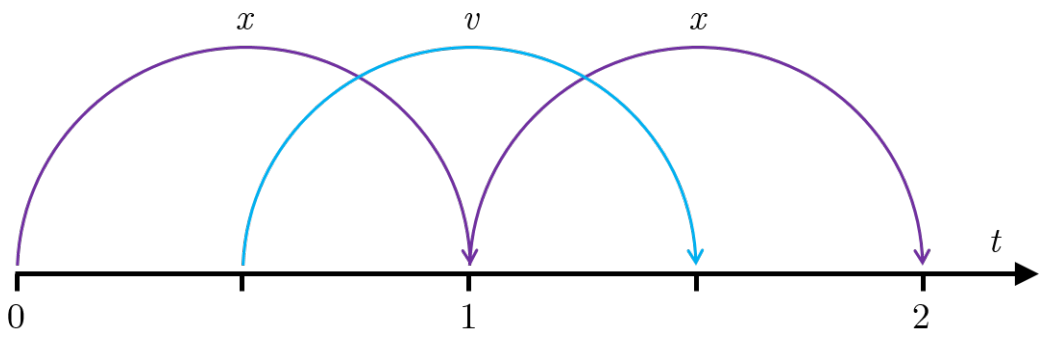


Figure 2.8: Representation of the leap-frog integration method. The algorithm is called 'leap-frog' because x and v are leaping over one another.

due to high energy overlaps between atoms. With an appropriate time step, the phase space is covered efficiently and collisions between atoms occur smoothly.[96]

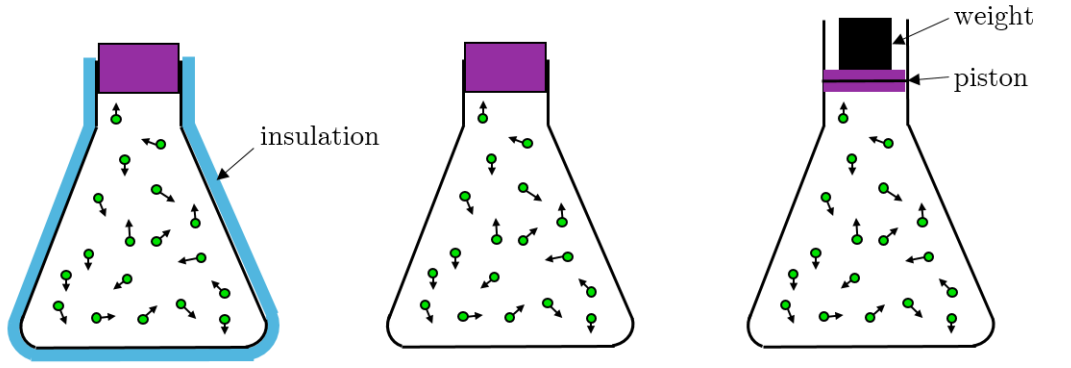
2.4.2 Ensembles

In the early days of MD, the microcanonical (NVE; constant number of particles N, volume V and total energy E; Fig. 2.9a) ensemble was traditionally used for MD simulations. In such an ensemble, the system remains isolated and unable to exchange energy or particles within its environment.[111] As a result, these conditions do not closely represent experimental environments, which are typically temperature controlled. Instead, it is more appropriate to maintain a constant temperature T and/or pressure P throughout the simulation, to match more closely to the experimental conditions. For this purpose, more recently the canonical (NVT) and isothermal-isobaric (NPT) ensembles (Fig. 2.9b and c) are commonly used. In order to run MD simulations in these ensembles, a thermostat or barostat is necessary to maintain the constant temperature or pressure of the system.

2.4.2.1 Thermostats

There are several different thermostats available, which allow a temperature to fluctuate around an average by adding or removing energy to and from the system in a constant temperature ensemble. The most commonly used thermostats in MD simulations of biological molecules are the Berendsen[112] and Nosé-Hoover[113] thermostats, both of which were used in this study.

The Berendsen thermostat works by weakly coupling the system to a thermal bath with a coupling constant, τ_T, and rescaling the particle velocities proportional to the differences between the system temperature and the temperature of the thermal bath at each time step. The larger the value of τ_T, the longer the system



a) Microcanonical (NVE) b) Canonical (NVT) c) Isobaric-isothermal (NPT)

Figure 2.9: Representation of (a) the microcanonical (constant NVE), (b) the canonical (constant NVT), and (c) the isobaric-isothermal (constant NPT) ensembles.

will take to adjust to the target temperature, T_0, from the temperature at time $T(t)$:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T} \tag{2.25}$$

The Berendsen thermostat has the advantage that it is relatively simple to implement and that the value of τ_T can be easily changed to suit the requirements of the system. However, the Berendsen thermostat suppresses the fluctuations in kinetic energy, so the system never approximates a true canonical ensemble. The error in ensemble averages are small for large systems, but are still sufficient to affect the distribution of kinetic energy. Therefore, any properties calculated from the fluctuation of the kinetic energy will be incorrect. Depending on the system, the Berendsen thermostat can be used when computing some average properties. In this work, where the Berendsen (or its adaptation, velocity rescaling[114]) thermostat was used, we did not calculate any properties derived from the kinetic energy.

The Nosé-Hoover thermostat is implemented by introducing an extra degree of freedom into the Hamiltonian of the system to take the form:

$$H(P, R, p_s, s) = \sum_i \frac{P_i^2}{2ms^2} + \frac{1}{2}\sum_{i,j,i\neq j} U(R_i - R_j) + \frac{p_s^2}{2Q} + gk_bT\ln(s) \tag{2.26}$$

where P and R represent all co-ordinates r_i and p_i, Q is a coupling constant, p_s is the momentum of the extra degree of freedom, s, and g is the number of independent momentum degrees of freedom.

The Nosé-Hoover thermostat is able to approximate a canonical ensemble, unlike the Berendsen thermostat, which means it can be used to calculate properties

derived from the fluctuation of the kinetic energy e.g. the heat capacity. Additionally, while the Berendsen thermostat is very efficient for relaxing a system to the target temperature, the Nosé-Hoover thermostat is also computationally inexpensive.

2.4.2.2 Barostats

Several different barostats exist, which maintain constant pressure of a system during a simulation in the isothermal-isobaric (NPT) ensemble. Commonly used are the Berendsen[112] and Parrinello-Rahman[115] barostats.

Similar to the Berendsen thermostat, the Berendsen barostat controls the pressure by weakly coupling the system to a pressure bath with a coupling constant τ_P. The larger the value of τ_P, the longer the system takes to adjust to the target pressure, P_0, from the pressure at time $P(t)$:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \tag{2.27}$$

The Berendsen barostat does produce the correct average pressure, but it is not able to yield a true NPT ensemble.[112] In cases where the fluctuations in the pressure or volume are important (for example, if thermodynamic properties will be calculated from them), the Berendsen barostat may not be appropriate. Instead, the Parrinello-Rahman barostat may be used, often combined with the Nosé-Hoover thermostat, as in theory this yields a true NPT ensemble.

The Parrinello-Rahman is similar to the Nosé-Hoover thermostat in the way it uses an additional degree of freedom. This extended ensemble algorithm allows the volume and shape of the simulation cell to fluctuate (if the system is far from equilibrium the box may oscillate wildly). The Hamiltonian includes an additional term to represent a thermal reservoir s:

$$H = K + V + K_s + V_s \tag{2.28}$$

where K is the kinetic energy and V is the potential energy. The equations of motion for the atoms obtained from the Hamiltonian becomes:

$$\frac{d^2 r_i}{dt^2} = \frac{m_i}{F_i} - \xi \frac{dr}{dt} \tag{2.29}$$

where the acceleration of an atom i is reduced by some factor $\xi \frac{dr}{dt}$ (where ξ represents a fictitious friction parameter).

2.4.3 Constraints

As alluded to previously, choosing an appropriate time step is very important in order to ensure the simulation runs smoothly. The time step should ideally be small enough to capture all intramolecular atomic interactions, including bond vibrations. However, this would be extremely computationally expensive and slow to implement. Instead, constraints can be placed on bonds to eliminate vibrations, fixing the bond length to a set value during the simulation. This allows the time step to be increased without compromising on accuracy. The LINCS (Linear Constraint Solver) algorithm is a non-iterative method which resets bonds to their correct length after each integration step. Other algorithms such as SHAKE and SETTLE are implemented in GROMACS, but LINCS is predominantly used throughout this work, as it is three to four times faster than the SHAKE algorithm while maintaining accuracy.[116,117] In this work, all bonds were constrained using the LINCS algorithm.

Constraints are not to be confused with restraints. In this work, harmonic position restraints were placed on molecules during equilibration, preventing deviation from the starting position. Restrained molecules are able to move, but receive a large energy penalty for deviating from their starting position:

$$V_{pi}(r_i) = \frac{k_{pr}}{2}|r_i - r_0|^2 \tag{2.30}$$

where $V_{pi}(r_i)$ is the energy penalty imposed on particle i for moving to position r_i from starting position r_0.

2.4.4 Periodic Boundary Conditions

In order to calculate macroscopic properties of a system, the system size needs to be sufficiently large, which is challenging and computationally expensive to simulate. Periodic boundary conditions (PBC) enable a simulation to be performed using a relatively small number of particles, where the particles experience forces as if they were in a bulk phase.[96] The simulation cell is replicated in all directions to form an infinite lattice of identical cells, so a particle that moves out of the unit cell is replaced by an image particle that enters from the opposite side (Fig. 2.10), thus keeping the number of particles within the unit cell constant, and removing any effects due to boundaries to allow for sampling of bulk macroscopic effects. Each atom interacts with any neighbouring atom or image in the periodic lattice within the interaction cut-off specified by the force field. Therefore, it is important that the unit cell is large enough to accommodate the cut-off and prevent an atom from interacting with its own periodic image.

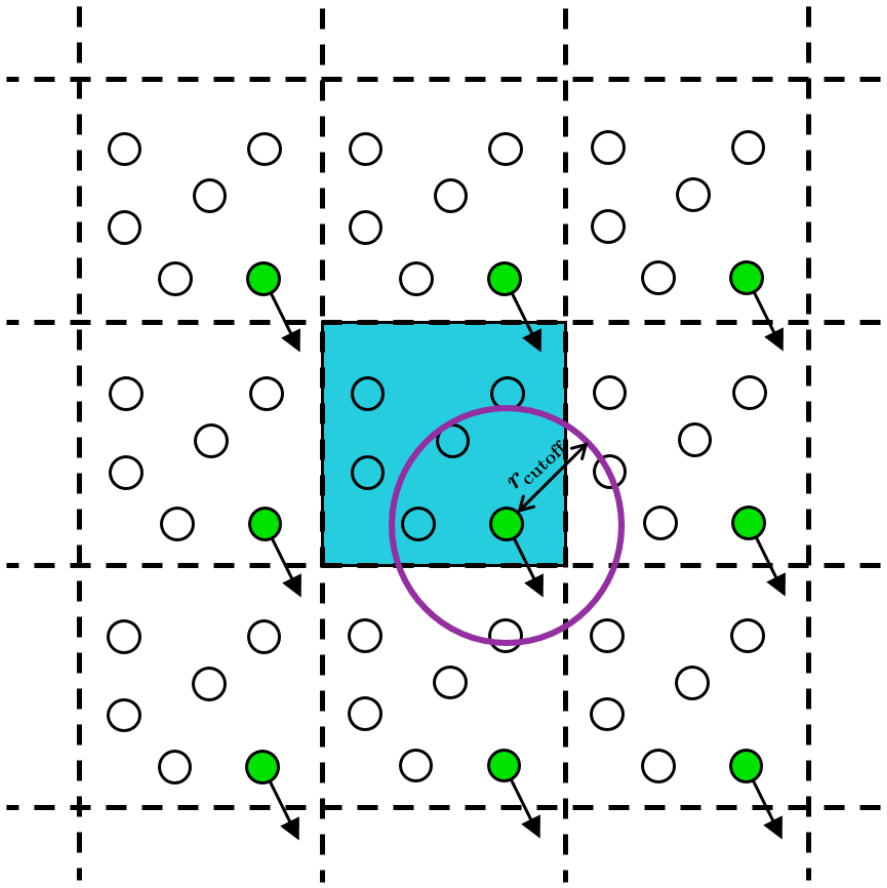


Figure 2.10: Representation of periodic boundary conditions in two dimensions. The simulation cell (blue) is replicated to form an infinite lattice. The particle experiences short-range interactions with all particles within the specified cut-off, r_{cutoff}.

2.4.5 Steered Molecular Dynamics

Many biologically relevant events occur on time scales that are inaccessible by conventional MD. Steered molecular dynamics (SMD) simulations can be used to bias the behaviour of a system towards a particular phenomenon, such as extraction of lipids from membranes, PPIs and protein-ligand binding.[118–120] Experimental techniques such as atomic force microscopy are based on the application of mechanical forces to single molecules and are commonly used to study the properties of biomolecules. In SMD simulations, time-dependent external forces are applied, for example by anchoring a molecule at a fixed point and pulling at another point via a dummy spring (Fig. 2.11), in a manner analogous to single molecule spectroscopy. It has previously been used to aid unwinding of helical structures,[121,122] as it allows unwinding events to occur on computationally feasible time scales. There are two typical ways to perform SMD simulations: constant force and constant velocity. In constant velocity SMD simulations (as performed in this work), a dummy spring (harmonic potential) is used to induce motion, and the free end of the spring is moved at constant velocity while the atoms attached to the fixed end of the spring are subject to the steering force. In constant force SMD simulations, a force is directly applied to one or more atoms, and the dynamics of the system are used to measure extension or displacement.[123]

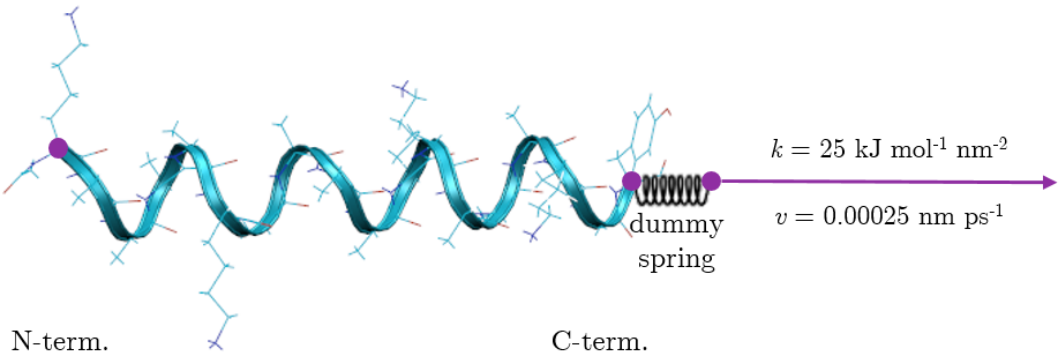


Figure 2.11: Representation of a constant velocity SMD simulation. Peptides are anchored at one point (represented by the N-terminal purple dot) and pulled by a dummy spring fixed at a second point. The force constant of the spring, k, and the pull speed, v can be optimised based on the system.

Qualitative data can be readily obtained from SMD simulations, and with sufficient repeats quantitative data can also be obtained. Jarzynski's principle states that thermodynamic potentials can be obtained from irreversible processes such as a protein unfolding. The ensemble average of an exponential of the total work W performed on the system during a non-equilibrium transition from one

state to another is connected to the free energy different ΔF of the two states[118]:

$$\langle exp[-W/k_BT]\rangle = exp[-\Delta F/k_BT] \tag{2.31}$$

In this work, SMD simulations were used to unwind parent and modified 18-residue helical peptides. These peptides were shown based on experimental data to have reduced helical content, and observing a reduction in helicity was not feasible using conventional MD.[93] The total amount of work required to pull each peptide 5 nm was calculated by integrating the area under each force extension curve produced from the SMD simulations, using the trapezium rule via a Python 3 script (included in Appendices).

2.5 Biophysics Theory

2.5.1 Circular Dichroism

Circular dichroism (CD) spectroscopy is a technique commonly used to investigate the secondary structure of proteins and peptides. Chiral molecules give CD signals at wavelengths where they absorb light. A molecule is considered chiral if it cannot be superimposed on its mirror image (Fig. 2.12).[124]

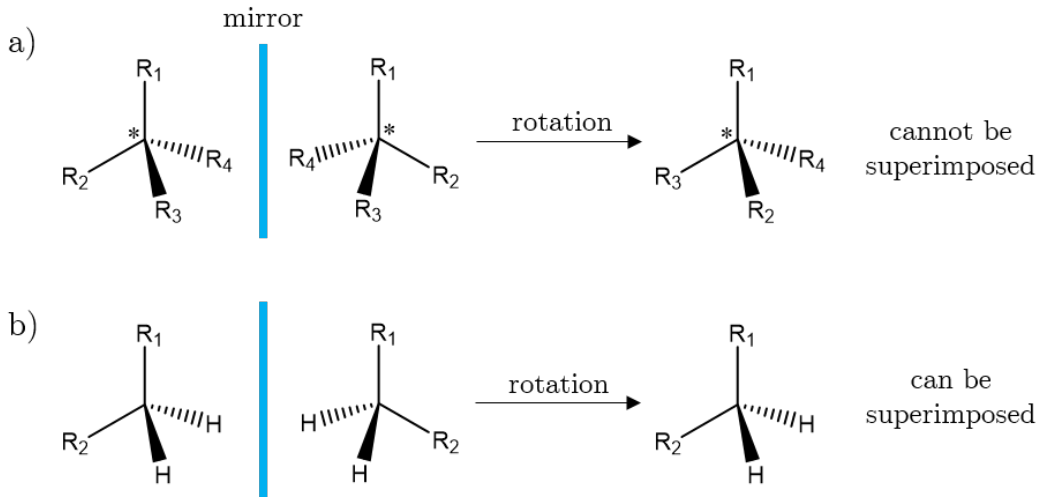


Figure 2.12: Comparison of (a) chiral and (b) achiral molecules. The chiral stereocentre is indicated with (*). Rotation of the mirror image of (a) does not general the original structure, in order to superimpose the mirror images bonds would have to be broken and reformed. Rotation of the mirror image of (b) can be superimposed on the original structure.

CD spectroscopy measures the difference in absorbance of left- and right-

handed circularly polarised light (A_L and A_R respectively) at a given wavelength:

$$\Delta A = \Delta A_L - \Delta A_R \tag{2.32}$$

This difference in absorption is normally outputted in units of millidegrees (θ). This can be normalised with respect to sample concentration, peptide size and cuvette path length by converting from millidegrees to mean residue ellipticity (MRE; Θ):

$$\Theta = \frac{\theta \times 0.1 \times M_r}{c \times l \times N_A} \tag{2.33}$$

where M_r is the mean residue weight, c is the concentration in $\mathrm{mg\,mL^{-1}}$, l is the path length in cm and N_A is the number of residues in the peptide. MRE is given in $\mathrm{mdeg\,cm^2\,dmol^{-1}}$.

Different structural elements have characteristic CD spectra (Fig. 2.13). For example, the CD spectrum for an α-helical peptide or protein has two negative maxima at 208 and 222 nm. Data can be fit using tools such as DichroWeb[125] to obtain an estimate of global secondary structure content.

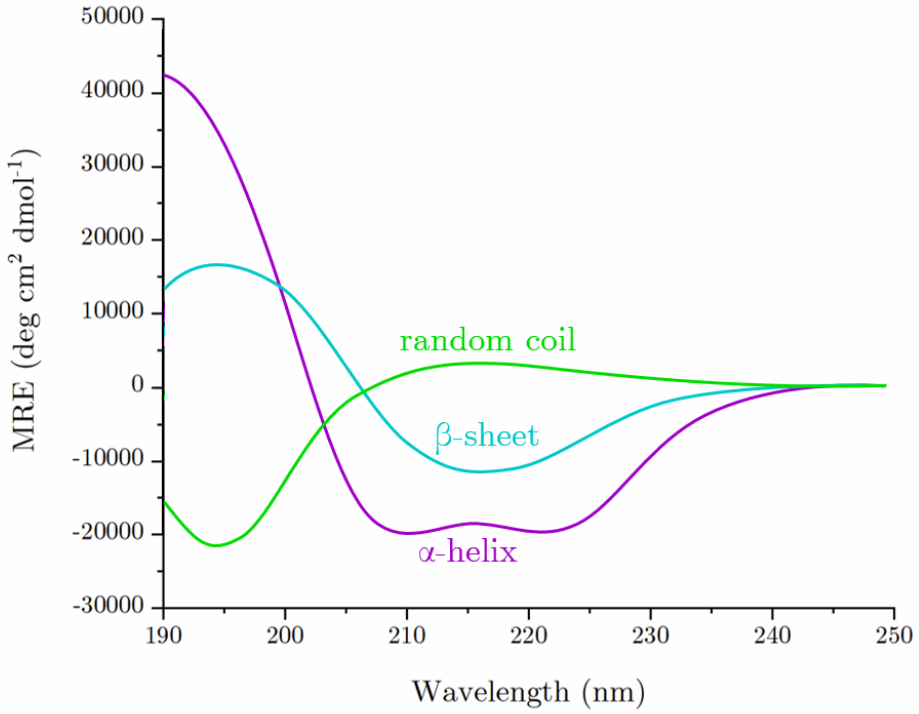


Figure 2.13: Characteristic spectra for α-helices (purple), β-sheets (cyan) and random coil (green) structures.

2.5.2 Solution-State NMR Spectroscopy

Solution-state NMR is a high resolution technique in which the chemical environments of individual atoms within a molecule can be studied. It is well-suited for the structural characterisation of peptides and small proteins. The basic principles of NMR are described briefly here.

2.5.2.1 Spin Angular Momentum

Electrons possess an intrinsic quantum mechanical property called 'spin'. Protons and neutrons within atomic nuclei also possess a spin angular momentum called the nuclear spin angular momentum. The spins of protons and neutrons contribute to the net spin of the atomic nucleus. If the spins are paired (for example in ^{12}C, $S = 0$), the atom is not NMR-active. If, however, the spins are unpaired, (for example in ^{1}H, $S = \frac{1}{2}$), then the atom possesses a net spin, and is able to be studied using NMR. Spin in atoms with unpaired nucleons (such as ^{1}H, ^{13}C and ^{15}N) produces a magnetic dipole moment, μ:

$$\mu = \gamma\hbar\sqrt{S(S+1)} \tag{2.34}$$

where $\hbar$ is Planck's constant divided by 2π, and γ is the gyromagnetic ratio. Since the value of γ is characteristic to each nuclei, as are the strengths of the magnetic dipole moments and therefore their NMR frequencies.

In the presence of a magnetic field B_0, for a sample containing nuclei of $S = \frac{1}{2}$ the spin of each nucleus aligns either with or against the magnetic field. Thus, $S = \frac{1}{2}$ nuclei will split into two different populations of energy levels, one parallel and the other one anti-parallel to B_0. This splitting is dependent on the chemical environment of each nucleus, and the ratio of populations aligning with or against the field is described by the Boltzmann distribution. At thermal equilibrium, there is a small excess number of nuclei in the lower energy state (i.e. aligned with B_0), resulting in a small but measurable net magnetisation, M_0.

2.5.2.2 The Larmor Frequency

When a particle is placed in an external magnetic field B_0, the magnetic moment precesses around the z-axis at a frequency known as the Larmor Frequency, ω_0 (Fig. 2.14). The Larmor frequency depends on the gyromagnetic ratio, γ, and the magnetic field applied, B_0:

$$\omega_0 = -\gamma B_0 \tag{2.35}$$

NMR spectroscopy measures the magnetisation perpendicular to the z-axis.

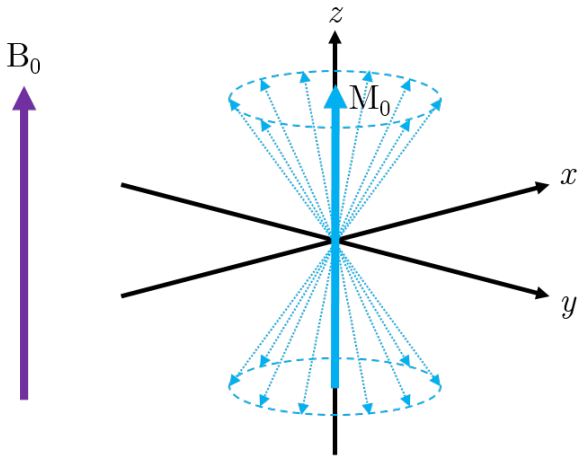


Figure 2.14: Precession of spin around magnetic field B_0, indicated by the dashed circles. Since there is a slight excess of spin aligned to B_0, there is a net magnetisation M_0 aligned along the z-axis.

When M_0 is aligned along the z-axis, an NMR signal is not generated. In order to generate an NMR signal, M_0 must be perturbed away from the z-axis. Application of a radio frequency (RF) pulse along the x- or y-axis causes the bulk magnetisation to precess in the xy-plane at a frequency related to the RF field (Fig. 2.15). This generates an electrical current which is detected by a receiver coil within the NMR spectrometer.

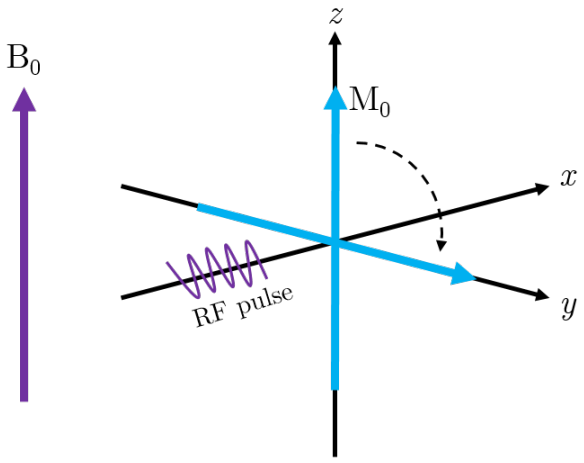


Figure 2.15: Application of a radio frequency pulse perpendicular to the external field B_0 moves the net magnetisation M_0 into the xy-plane.

It is possible to alter the angle of rotation of the magnetisation by varying the length of time for which the RF field is applied. The most common pulse is the 90° excitation pulse. This pulse causes the spins to cluster, generating a transverse magnetisation.

2.5.2.3 Free Induction Decay

After application of the RF pulse, the magnetic moment relaxes back to its thermal equilibrium along the z-axis. The rate at which this occurs depends on T_1 and T_2, known as the longitudinal and transverse relaxation rates respectively. T_1 describes the return of magnetisation to equilibrium as a result of spin-lattice interaction, and T_2 describes the return to equilibrium as a result of spin-spin interactions. As the magnetisation relaxes, precession continues, and the radius of precession shortens as M_0 approaches the z-axis. As a result, following termination of the RF pulse, the amplitude of the detected NMR signal decays over time. The decay and oscillation of the transverse magnetisation over time can be measured and is called the free induction decay (FID) (Fig. 2.16a and b). Fourier transformation of the FID yields a frequency spectrum (Fig. 2.16c).

2.5.2.4 Nuclear Shielding

Each NMR-active nucleus within the magnetic field acts as a tiny electromagnet. This means that each nucleus imposes its own local magnetic field, thus a small magnetic field B' is established in either the opposite direction or aligned to the applied field B_0. As a result, the nucleus is shielded or deshielded from the external magnetic field by surrounding electrons and consequently experiences an altered total magnetic field as shown in Eqn. 2.36:

$$B = B_0 - B' = B_0(1 - \sigma) \tag{2.36}$$

where σ is the shielding constant.

Electronegative atoms such as oxygen are electron-withdrawing groups, and as a result the nuclei within are deshielded and experience a greater level of magnetisation than expected, precessing with corresponding higher frequencies (signal shifted downfield in the spectrum, Fig. 2.17). Conversely, nuclei within electropositive groups are shielded from the magnetic field, and so the frequencies of these nuclei are reduced (upfield in the spectrum).

In molecules, electron currents provided by neighbouring groups can also lead to shielding or deshielding of nuclei. Therefore, the frequency of a nucleus depends also on its local chemical environment. Different nuclei in a molecule resonate at slightly different frequencies causing each nucleus to precess at a different Larmor frequency, thereby creating different resonance frequencies (chemical shift) in the NMR spectrum. Frequencies are described with reference to the frequency of a standard, typically tetramethylsilane (TMS) when protons are studied in aqueous solution. TMS contains 12 hydrogen atoms which are all in the exact same chemical

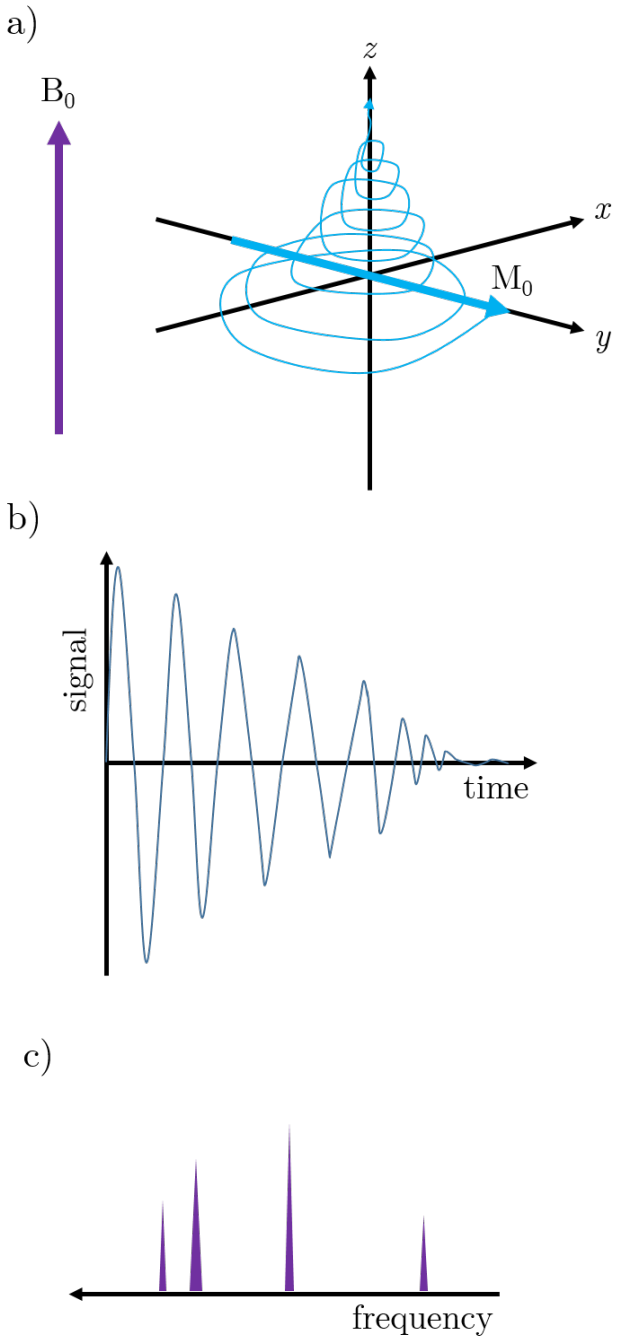


Figure 2.16: (a) The precession and relaxation of the net magnetisation M_0 back to longitudinal magnetisation. (b) The signal is detected as function of time along the xy-plane (FID). (c) The signal is Fourier transformed to obtain the frequency spectrum.

environment, resulting in a strong single peak. These methyl hydrogens are highly shielded from the external magnetic field, so adopt very upfield Larmor frequencies.

Frequencies of studied protons are referenced with respect to that of the stan-

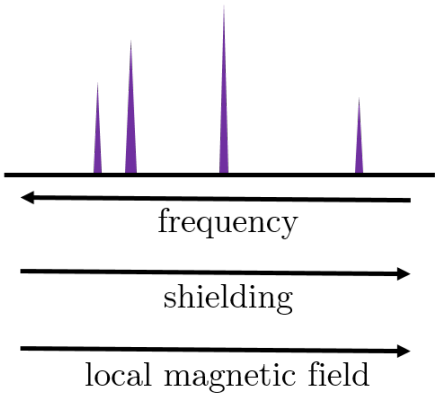


Figure 2.17: The 1D ^{1}H NMR spectrum showing where more or less shielded nuclei appear on the spectra.

dard frequency, and with respect to the field strength of the NMR spectrometer:

$$\delta_{ppm} = \frac{\sigma_0 - \sigma_{ref}}{\sigma_{ref}} \times 10^6 \tag{2.37}$$

where σ_{ref} is the Larmor frequency of the methyl protons in the standard.

The chemical shifts of nuclei are specific to a particular functional group (such as aliphatic protons) and are descriptive of the chemical structures within the sample. A disperse set of chemical shifts indicates that nuclei within the sample experience a number of different chemical environments. Such a spectrum might indicate, for example, that a protein of interest is well folded. In an unstructured or aggregated protein, the peaks are broader and not as widely dispersed due to loss of different local environments.

2.5.3 2D NMR Techniques

For large biomolecules such as proteins, it is often be impossible to distinguish between NMR signals arising from each nucleus in a one-dimensional ^{1}H spectrum, as signals overlap, especially where there are several nuclei in similar chemical environments. To overcome problems caused by overlap in the 1D spectrum, two-dimensional NMR spectroscopy can be used.

A 1D NMR spectrum has two dimensions: the x-axis corresponds to the chemical shifts in ppm, and the y-axis corresponds to the intensity (Fig. 2.18a). A 2D NMR spectrum contains two frequency axes, where the horizontal axis is for the direct dimension (F2) and the vertical axis is the indirect dimension (F1) with intensity as the third axis. As a result, 2D NMR spectra are typically displayed as contour plots (Fig. 2.18b).

In homonuclear spectra (e.g. both proton frequencies in the two frequency

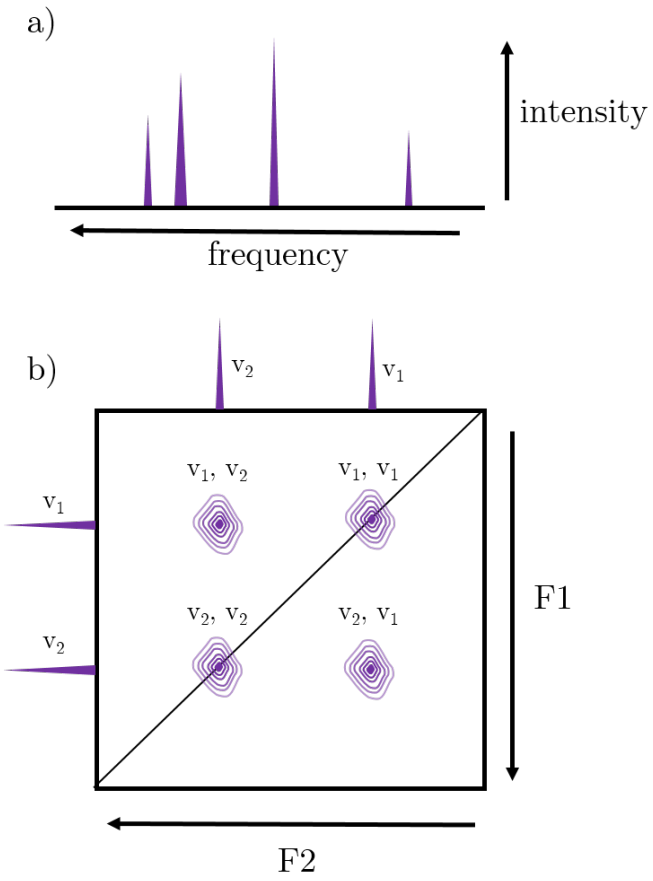


Figure 2.18: Representations of (a) a 1D NMR spectrum, with two dimensions corresponding to frequency and intensity, and (b) a 2D NMR contour plot, with two frequency dimensions (F1 and F2) and intensity as the third axis. A homonuclear spectum is symmetrical and contains cross peaks (v_1, v_2; v_2, v_1) and diagonal peaks (v_1, v_1; v_2, v_2).

dimensions), peaks are symmetric with respect to the diagonal of the spectrum. The diagonal peaks correlate identical spins, and are of little analytical use.

A 2D NMR pulse program has a general basic format composed of four periods: preparation, evolution, mixing and detection (Fig. 2.19). The preparation and mixing periods vary depending on the nature of the experiment, and the detection period is analogous to the detection period of a 1D experiment. It is the evolution period, t_1, that allows the generation of a second (indirect) dimension. During the recording of a 2D experiment, the same experiment is repeated many times with a varying t_1. The repeated acquisition of FIDs with systematically incremented t_1 time periods, followed by a double Fourier transform with respect to t_2 and t_1 results in the 2D spectrum.[126]

2.5.3.1 Total Correlation Spectroscopy

Total correlation spectroscopy (TOCSY, also called HOHAHA (Homonuclear Hartmann Hahn),[127,128] provides through-bond proton-proton correlations due to scalar couplings (also known as J-couplings, Fig. 2.20a) between all the atoms in a single spin-system, i.e. all the protons in a single amino acid (Fig. 2.20b). Scalar coupling arises due to indirect interaction between two nuclear spins, mediated by

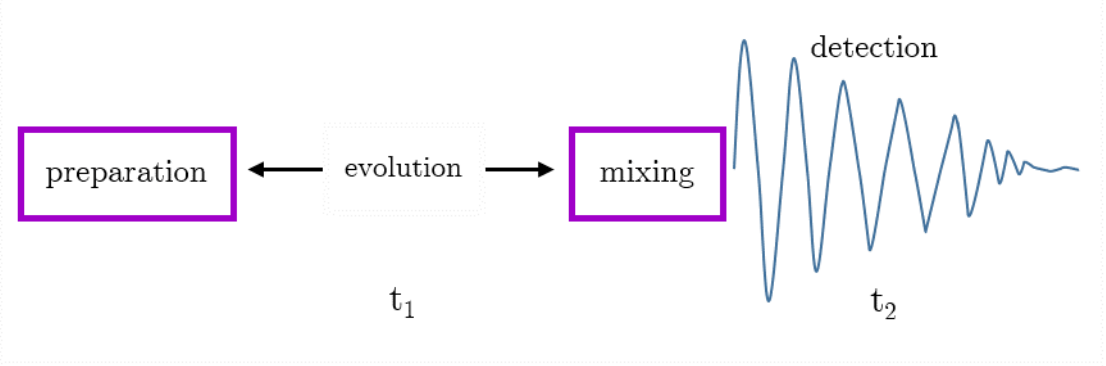


Figure 2.19: The general scheme for any 2D experiment. Reproduced from Claridge.[126]

the electrons participating in the bond connecting the nuclei. Information cannot travel from amino acid to the next because the carbonyl carbon of the peptide backbone cannot participate in scalar coupling in ^{1}H-^{1}H NMR. Each amino acid has a characteristic pattern of chemical shifts which can be used to identify the amino acid (Fig. 2.20c).

2.5.3.2 Nuclear Overhauser Enhancement Spectroscopy

The NOESY (nuclear Overhauser enhancement spectroscopy) experiment is crucial for the determination of peptide structure. It uses through-space proton-proton correlations due to dipolar couplings (the nuclear Overhauser effect, NOE) (Fig. 2.21a). The NOE describes the transfer of nuclear spin polarisation from one nucleus to another via crossrelaxation.[130] It identifies the chemical shifts of pairs of protons that are close together in three-dimensional space, regardless of whether or not they are within the same amino acid. The coupling between two protons depends on the distance between them (r_{ij}), but a signal is only observed if they are within 5 Å of each other. The rate of build-up of the NOE is proportional to ${r_{ij}}^{-6}$ (Fig. 2.22),[131] and thus the size of the NOE is a sensitive measure of inter-nuclear distance:

$$\sigma_{ij} = \frac{3}{10}\frac{\gamma^4\hbar^2}{{r_{ij}}^6}\left[\frac{2\tau_c}{1+4\omega^2\tau_c^2} - \frac{\tau_c}{3}\right] \tag{2.38}$$

where r_{ij} is the inter-nuclear distance between nuclei i and j, σ_{ij} is the rate of growth of the NOE, τ_c is the correlation time, ω is the resonance frequency for ^{1}H and γ is the magnetogyric ratio for ^{1}H.

The sign of the NOE depends on the magnitude of the correlation time (τ_c; the time it takes for the compound to rotate through 1 radian)[132] and thus the molecular weight. For a small peptide, cross-relaxation occurs slowly, and NOEs

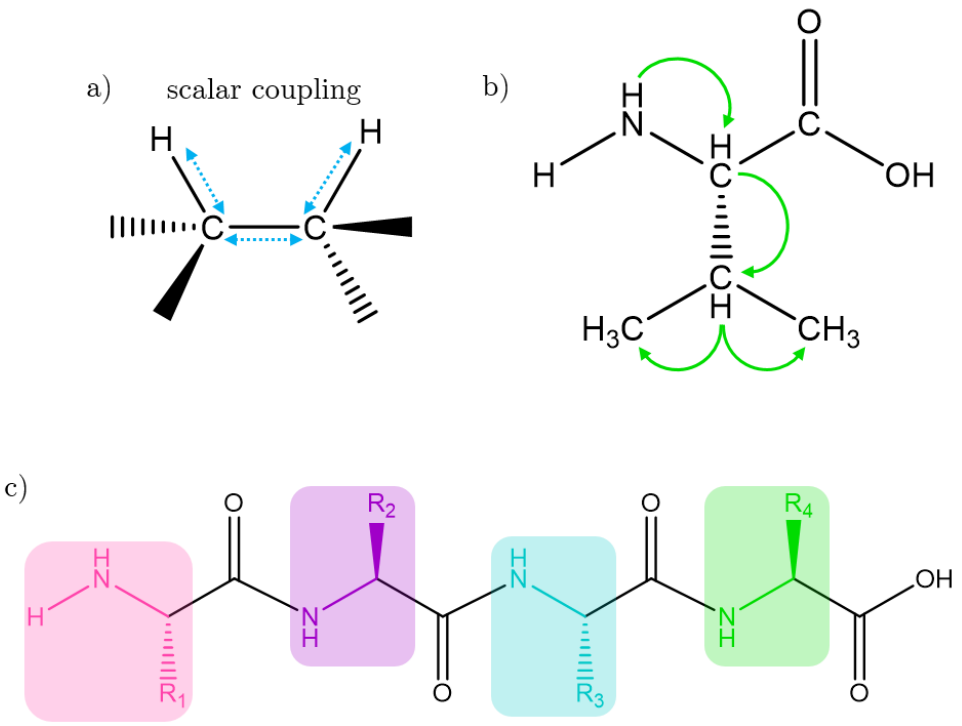


Figure 2.20: Information provided by a TOCSY experiment. (a) TOCSY cross-peaks arise from scalar coupling, and (b) the magnetisation can be transferred over up to 5 or 6 bonds as successive protons are coupled. (c) The carbonyl group disrupts the TOCSY transfer, thus TOCSY only allows identification of all protons within a single spin system. Adapted from Roberts.[129]

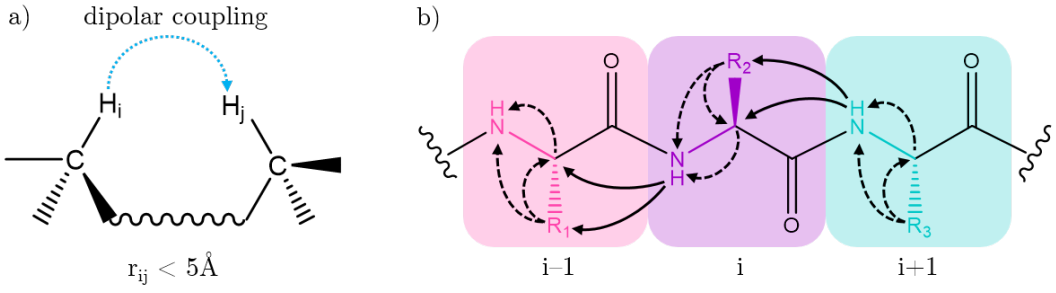


Figure 2.21: Information provided by a NOESY experiment. (a) NOESY cross-peaks arise from dipolar coupling only for nuclei within 5 Å of each other. (b) NOESY provides intra-residue correlations (dashed lines) as well as inter-residue correlations (solid lines), which allow the peptide to be assigned sequentially.

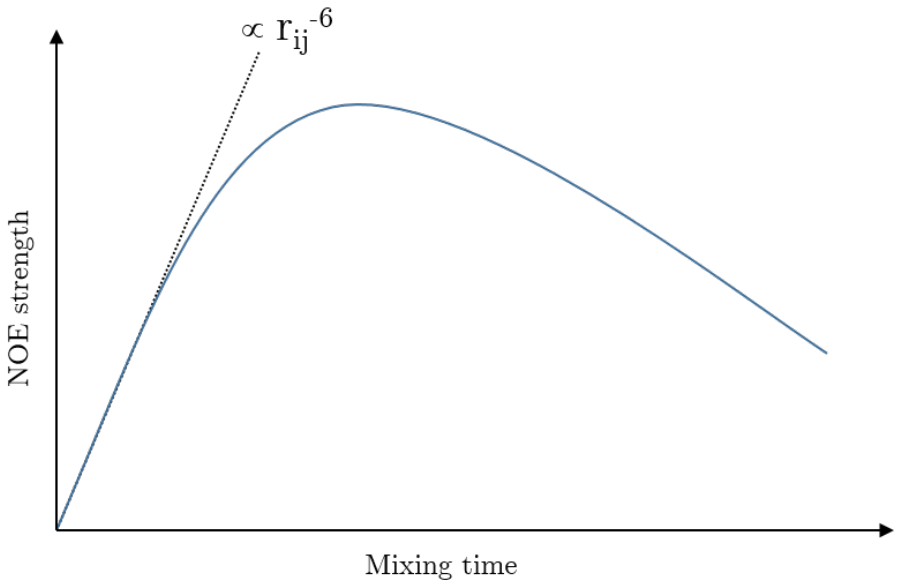


Figure 2.22: NOE strength is dependent on time. At low mixing times, the NOE builds up quickly at a rate that is proportional to $r_{ij}{}^{-6}$. As mixing time increases, the NOE begins to decay due to relaxation. Adapted from Claridge.[126]

build up steadily. As the molecular weight of the sample increases, cross-relaxation gets faster, and NOEs build up more rapidly. Fig. 2.23 represents the maximum proton-proton NOE as a function of molecular tumbling rates. The curve can be divided into three distinct regions, which correspond to fast, intermediate and slow motion regimes. For low molecular weight molecules (short τ_c) that tumble rapidly, the NOE is positive. For molecules that are very large and tumble very slowly, the NOE is negative. In between these regimes, the NOE passes through zero and then becomes negative, thus for peptides with molecular weights of ∼1500 Da, the NOE may become very small or even zero, and alternative techniques must be used. In such a situation, ROESY (rotating-frame Overhauser enhancement spectroscopy) can be used instead of NOESY. In ROESY, the ROEs (rotating-frame Overhauser enhancements) are always positive regardless of correlation time.[126] Fortunately, for all the peptides used in this study, NOESY proved to be a suitable technique and ROESY was not necessary.

Because the NOESY can contain both inter-residue and intra-residue correlations, it is possible to deduce which residues are next to each other in the peptide chain (Fig. 2.21b). As it correlates protons which are distant in the amino acid sequence, but are close in space due to higher order structure, it allows for the identification of secondary and tertiary structure elements.

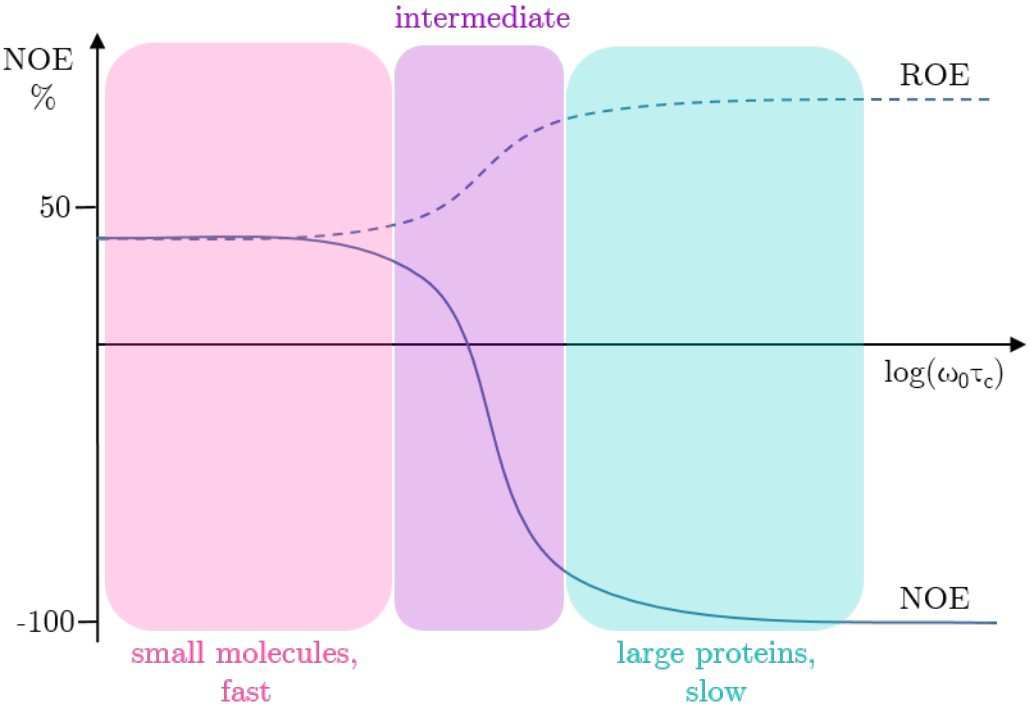


Figure 2.23: NOE and ROE as a function of molecular tumbling rate, $\omega_0\tau_c$, where ω_0 is spectrometer observation frequency and τ_c is the correlation time. Adapted from Claridge.[126]

2.5.3.3 Heteronuclear Techniques

For large and complex molecules such as proteins, ^{1}H NMR signals can become very challenging to resolve. Whilst two-dimensional techniques such as TOCSY can solve this, heteronuclear techniques such as heteronuclear single-quantum correlation (HSQC) can also help overcome this constraint. In an HSQC spectrum, two types of nuclei are excited and mixed, and the two frequency axes correspond to the two different isotopes, with protons along the F2 axis and a heteronucleus (such as ^{13}C or ^{15}N) along the F1 axis. Each peak is due to coupling between ^{1}Hs and heteronuclei across a single bond. As such, it identifies directly connected nuclei.[126] For example, a ^{1}H-^{13}C HSQC spectrum will display all ^{1}H-^{13}C correlations within the target protein, so can provide structural information about amino acid sidechains. Similarly, a ^{1}H-^{15}N HSQC experiment will provide information about all ^{1}H-^{15}N correlations, such as those present in the backbone amides of an amino acid, as well as in some amino acid sidechains. The spectral window for ^{13}C and ^{15}N are much wider than for ^{1}H, therefore HQSC spectra can assist in the assignment of a molecule by providing additional information about the chemical environment of an atom in a manner which is easy to distinguish (signal overlap is greatly reduced).

The major disadvantage of heteronuclear techniques is that they require the presence of NMR-active nuclei other than ^{1}H, which can be low in abundance. Whilst natural-abundance can be sufficient to obtain HSQC spectra for small molecules (in our case it was sufficient for the small peptides), for larger molecules such as proteins it may be necessary for the molecule to be isotopically enriched with ^{13}C or ^{15}N.

Chapter 3

Methodology

3.1 Peptide Preparation

Synthetic peptides were prepared in-house or by Insight Biotechnology Ltd. (Middlesex, UK) according to Table 3.1. Sequences for each peptide are listed in Table 3.2.

Table 3.1: Peptides used throughout this work. All peptides were made in-house (indicated with *) or by Insight Biotechnology Ltd., Middlesex, UK.

Sample	Source	Notes
LAGAY derivatives	Dr Stefan Roesner*	Synthesis and characterisation previously described[92]
Helical peptides	Dr Jonathan Beadle*	Synthesis and characterisation previously described[10]
Bacaucin-1 derivatives	Dr Ina Wilkening*	
Panel of AMPs (Table 3.6)	Insight Biotechnology Ltd.	
Ac-RRWWCA-NH_2 derivatives	Dr Sam Spring*	

3.2 Circular Dichroism

CD spectra were recorded using a Jasco J-1500 spectropolarimeter (Jasco UK, Essex, UK), using acquisition parameters listed in Table 3.3. Peptide samples for CD were prepared to be approximately 175 μM in 10 mM potassium phosphate buffer (pH 7.0), and spectra were acquired using a quartz cuvette with a path length of 0.1 cm.

For all samples, baseline spectra were acquired of the buffer solution. The baseline spectra were subtracted from peptide spectra prior to conversion from

Table 3.2: Sequences for all peptides used throughout this work.

Samples	Sequence
LAGAY derivatives	LAGAY
	AGAYL
	$LAG_{ox}AY$
	$AG_{ox}AYL$
	LAGAY-OMe
	$LAG_{ox}AY$-OMe
	$LAG_{az}AY$-OMe
	cLAGAY
	$cLAG_{ox}AY$
Helical peptides	Ac-KAAAAKAAAAKAAAA-KGY-NH_2
	Ac-KAAAAKAA$_{ox}$AAKAAAA-KGY-NH_2
	Ac-KAA$_{ox}$AAKAAAAKAAAA-KGY-NH_2
	Ac-KAG$_{ox}$AAKAAAAKAAAA-KGY-NH_2
Bacaucin-1 derivatives	ELLSRVD
	ALLSRVD
Panel of AMPs	Ac-RRWWCA-NH_2
	GLLKRIK-NH_2
	KNKGWWW
	ALLRL
	KVFLGLK
Ac-RRWWCA-NH_2 derivatives	cRRWWCA
	Ac-RRWWGA-NH_2
	cRRWWGA
	$cRRWWG_{az}A$

millidegree to MRE.

3.3 Solution-state NMR Spectroscopy

3.3.1 ^{1}H 1D NMR Experiments

All NMR experiments were conducted at 283 or 298 K (for the helical and small peptides respectively) on a 700 MHz Bruker Avance spectrometer equipped with a cryoprobe, housed at the University of Warwick. 200 μL samples of 2–60 mM peptide were prepared in d_6-DMSO (small peptides) or 80:20 MeOD:H_2O (helical peptides). Spectra were acquired with 64k data points and 32 scans. Water suppression was achieved using excitation sculpting.[134] Data were processed using a Gaussian multiplication window function. Processing was performed using TopSpin 3.2.

Table 3.3: Acquisition parameters used for the Jasco J-1500 circular dichroism spectrometer. Reproduced from Lockey.[133]

Parameter	Value
Measurement range	180–300 nm
Data pitch	0.2 nm
CD scale	200 mdeg/1.0 dOD
FL scale	200 mdeg/1.0 dOD
Digital integration time	2 sec
Bandwidth	2.0 nm
Start mode	Immediately
Scanning speed	100 nm min^{-1}
Baseline correction	None
Shutter control	Auto
Accumulations	32

3.3.2 ^{1}H 2D NMR Experiments

200 μL samples of 2–60 mM peptide were prepared in d_6-DMSO (small peptides) or 80:20 MeOD:H_2O (helical peptides). Spectra were acquired with 256 data points in the F1 dimension and 4k data points in the F2 dimension, with 16 (small peptides) or 32 (helical peptides) scans. Water suppression was achieved using excitation sculpting[134] or the WATERGATE method.[135] The spectral width was 14 ppm in both dimensions for the cyclic and helical peptides, and 16 ppm in both dimensions for the short linear and methyl ester peptides. Data were zero-filled to 8k in F2 and 512 in F1 and processed using a squared sine window function and quadratic baseline corrections in both dimensions. Processing was performed using TopSpin 3.2, and analysis was performed using CcpNmr Analysis 2.4.2.[136]

The pulse programs and mixing times used in the TOCSY and NOESY experiments are summarised in Table 3.4.

Table 3.4: Pulse programs and mixing times used in TOCSY and NOESY experiments.

Sample	Pulse program	Mixing times (ms)
Short linear peptides	mlevphpr.2	70, 140
	noesyphpr	100, 250, 400, 600, 800
Methyl esters of peptides	mlevphpr.2	70, 140
	noesyphpr	100, 250, 400, 600, 800
Cyclic peptides	mlevphpr.2	70, 140
	noesyphpr	100, 250, 400, 600, 800
Helical peptides	dipsi2esgpph	70, 140
	noesygpphw5	100, 150, 200

3.3.3 HSQC Experiments

Natural-abundance HSQC experiments were used to assist with peptide assignment.

Two-dimensional ^{1}H-^{13}C HSQC spectra were obtained at 298 K using the pulse program hsqcfetgpsp.2, with 256 data points in the F1 dimension and 2k data points in the F2 dimension and 32 scans. The spectral width was 16 ppm in the F2 dimension and 260 ppm in the F1 dimension. Water suppression was achieved using z-gradients. Data were zero-filled to 1k in both dimensions, processed using a squared sine window function in both dimensions, and baseline corrected in the F2 dimension using a quadratic baseline correction.

Two-dimensional ^{1}H-^{15}N HSQC spectra were obtained at 298 K using the pulse program hsqcf3gpph19, with 256 data points in the F1 dimension and 4k data points in the F2 dimension and 32 scans. The spectral width was 16 ppm in the F2 dimension and 80 ppm in the F1 dimension. Water suppression was achieved using z-gradients. Data were zero-filled to 512 in F1 and 8k in F2, processed using a squared sine window function in both dimensions, and baseline corrected in the F2 dimension using a polynomial baseline correction.

Processing was performed using TopSpin 3.2, and analysis was performed using CcpNmr Analysis 2.4.2.[136]

3.4 Theoretical Methods

3.4.1 Force Field Parameters

Molecular dynamics (MD) simulations were performed using GROMACS 5.1.4[137] and the CHARMM27 force field[97,138] with additional parameters for the oxetane modification.[7] Starting configurations of the peptides were prepared using Avogadro 1.2.0.[139]

3.4.1.1 Oxetane Parameterisation

The CHARMM-compatible force field parameters for the 3-amino oxetane residue were derived by Powell et al.,[7] and a brief summary is described here.

Parameters were developed for small molecular fragments that capture the essential chemistry of the modified peptide bond (carbonyl to oxetane substitution and rotation about the ψ and ω dihedral angles). Non-bonded Lennard-Jones parameters for the C, O and H atoms comprising the oxetane ring were taken from existing CHARMM parameters for tetrahydrofuran (THF).[140]

Atomic partial charges were determined using the Force Field Toolkit (ffTK) Plugin[141] to the Visual Molecular Dynamics (VMD) software package,[142] and quantum mechanical (QM) calculations were performed using the Gaussian 03 program.[143] Firstly, a 3-amino oxetane molecule was energy minimized at the MP2/6-31G* level of theory. Putative hydrogen bond acceptor or donor interaction sites were identified and a TIP3P water molecule placed in an ideal orientation for hydrogen bonding with each interaction site. For each oxetane-water complex, calculations were carried out at the HF/6-31G* level of theory to optimize the distance of the water molecule to the target atom, and the interaction energy was determined from the difference between the energy of the complex and the independent energies of an oxetane and a water molecule. Target data for fitting bond stretching, angle bending and dihedral angle rotation parameters were obtained from QM calculations at the MP2/6-31G* level of theory.

3.4.2 Linear Methyl Ester Peptides

3.4.2.1 Model Building

The peptides were built assuming a linear conformation (ϕ, ψ and ω backbone dihedral angles were set to 180°) and uncharged ends. Both peptides had the C-terminus replaced with a methyl ester end group, and for the oxetane-modified peptide, the carbonyl of the glycine was substituted with an oxetane. This was followed by a steepest descent energy minimisation in Avogadro using the Universal Force Field,[144] which generated the starting structure of the two peptides for subsequent MD simulation in GROMACS 5.1.4.[137] Input topology files for GROMACS were generated using the GROMACS pdb2gmx tool after first modifying the residue database to include oxetane-substituted residues and the methyl ester end C-terminus. Each peptide was subjected to a short simulation (5000 steps) in the *NVT* ensemble in vacuum at 300 K, and was then solvated with a pre-equilibrated box of 424 or 426 DMSO molecules.[145] This box size was sufficient to ensure the peptides were not able to interact with their periodic image.

3.4.2.2 Simulation Parameters

Following solvation in DMSO, the system was subjected to 50000 steps of steepest descent energy minimization. This was followed by 100000 steps of simulation at 300 K in the *NVT* ensemble and 50000 steps of simulation at 300 K and 1 bar in the *NPT* ensemble to equilibrate the temperature and density of the system respectively. Each peptide was then simulated at a higher temperature of 500 K for 100 ns in the *NVT* ensemble, to enhance the sampling of conformational space

and prevent the conformations becoming trapped in local energy minima.

Cluster analysis (described in section 3.4.2.3) was performed on the trajectory to group peptide conformations based on their structural similarity. The central structure of the top five most populated clusters (which accounted for $> 99.9\%$ of the total population) was then used as the starting configuration for five independent 100 ns simulations of each peptide at 300 K and 1 bar in the *NPT* ensemble. As such, each peptide was simulated for a total simulation time of 500 ns.

All bonds were constrained using the LINCS algorithm[117] and a simulation time step of 2 fs was used. Periodic boundary conditions were applied in all directions. Lennard-Jones interactions were cut-off at 1.0 nm. Electrostatic interactions were handled using the particle mesh Ewald approach with a real-space cut-off of 1.0 nm. The temperature was controlled using velocity rescaling with a time constant of 0.1 ps,[114] and the pressure was isotropically maintained at 1 bar using the Parrinello-Rahman barostat with a time constant of 2.0 ps and compressibility of 4.5×10^{-5} bar^{-1}.[113,115] Atomic coordinates were saved every 20 ps for all analysis except the cluster analysis, which was performed on snapshots spaced 40 ps apart to avoid computer memory issues.

To improve the quality of the models, selected NOE distances from NMR experiments were incorporated as distance restraints in the MD simulations. The GROMACS implementation of distance restraints follows that of Torda et al., whereby time-averaged restraints are used.[146] Time-averaged distance restraints provide a better approximation of the physical nature of the NOE (which may reflect an averaging of multiple conformations), as they allow an atom to satisfy distance restraints that seem incompatible on average by moving between multiple positions.[137,146] If the NOE upper bound of the time-averaged distance between two atoms is exceeded, a harmonic restoring force (the strength of which is controlled by the corresponding force constant for the restraint and is described below) pulls the atoms back towards each other. The distance ranges and force constants were 1.8–2.7 Å and 2000 kJ mol^{-1} nm^{-2}, for strong restraints, 1.8–3.3 Å and 1500 kJ mol^{-1} nm^{-2} for medium restraints and 1.8–5.0 Å and 1000 kJ mol^{-1} nm^{-2} for weak restraints. The time constant for the distance restraints running average was 10 ps.

3.4.2.3 Convergence and Cluster Analysis

Cluster analysis was used to group structurally similar peptide conformations from the trajectory. We used the algorithm proposed by Daura et al. in which the root mean square deviation (RMSD) of atom positions between all pairs of structures is determined.[147] For each structure, the number of other structures for which the

RMSD of the backbone atoms was within 0.15 nm was calculated. The structure with the highest number of 'neighbours' was taken as the centre of a cluster and together with its neighbours was grouped as the first cluster. These structures were eliminated from the pool of structures, and the process was repeated to find the structure with the next highest number of neighbours to form the next cluster, until the pool of structures was empty. As structures were eliminated from the pool of structures once allocated to a cluster, each structure only belonged to one cluster.

In order to assess whether the number of clusters continues to change throughout the simulation, the number of clusters as a function of the total simulation time was calculated. A total of 4 clusters are identified for the parent peptide LAGAY-OMe, and no new structures are found after the third independent simulation (after approx. 250 ns total simulation time, Fig. 4.8a). A total of 5 clusters are identified for $LAG_{ox}AY$-OMe, and it takes five independent simulations (approx. 450 ns total simulation time) for the number of new structures to converge (Fig. 4.8b).

3.4.3 Cyclic Peptides

3.4.3.1 Model Building

The peptides were built assuming a linear conformation (ϕ, ψ and ω backbone dihedral angles were set to 180°) and uncharged ends. Both peptides were cyclised by manually joining the C-terminal carbonyl carbon to the N-terminal amide nitrogen with a long bond, and for the oxetane-modified peptide, the carbonyl of the glycine was substituted with an oxetane. The structures were briefly subjected to a steepest descent energy minimisation in Avogadro using the Universal Force Field,[144] which generated the starting structure of the two peptides for subsequent MD simulation in GROMACS 5.1.4.[137] This method of generating cyclic peptides is in line with other published methods,[148] and does not result in any significant variation of the backbones of the starting structures. Regardless, any deviation from the 'ideal' starting structure would likely be resolved during a 500 K *NVT* simulation as performed to enhance sampling. Input topology files for GROMACS were generated using the GROMACS pdb2gmx tool after first modifying the residue database to include oxetane-substituted residues and special bond information to appropriately model the cyclic backbone. As for the linear methyl ester peptides (section 3.4.2.1), each peptide was subjected to a short simulation (5000 steps) in the *NVT* ensemble in vacuum at 300 K, and was then solvated with a pre-equilibrated box of 425 or 428 DMSO molecules,[145] which was sufficient to ensure the peptides were not interacting with their periodic image.

3.4.3.2 Simulation Parameters

Simulation parameters were as described in section 3.4.2.2.

3.4.3.3 Convergence and Cluster Analysis

Cluster analysis and assessment of convergence was carried out as described in section 3.4.2.3, using a cut-off of 0.05 nm. This is a smaller cut-off than for the linear methyl ester peptides, which are much more flexible.

In order to assess whether our sampling was sufficient, the number of clusters as a function of the total simulation time was calculated. A total of 5 clusters are identified for the parent cyclic peptide, and no new structures are found after the third independent simulation (after approx. 250 ns total simulation time, Fig. 4.18a). Only 4 clusters are identified for the oxetane-modified cyclic peptide, and it also takes approx. 200 ns of total simulation time for the number of new structures to converge (Fig. 4.18b). Additionally, the cluster populations converge (Fig. 4.19) with the last 180 ns of simulation contributing no new structural information.

3.4.4 Helical Peptides

3.4.4.1 Brute Force Simulations

In order to compare how the model system behaved in different force fields, brief simulations using the GROMOS96 54a7, AMBER99SB-ILDN and CHARMM27 force fields were performed. The parent (unmodified) peptide was built in Avogadro 1.2.0[139] with ϕ and ψ angles corresponding to an idealised α-helix, with zwitterionic ends and protonated lysines. It was then briefly subjected to a steepest descent energy minimisation in Avogadro using the Universal Force Field.[144] The resulting structure was used for simulations in all three force fields using GROMACS 5.1.4.[137]

The initial structures were solvated in a box of 6942–6960 water molecules and charges were countered using chloride ions, and each system was subjected to 50000 steps of energy minimisation using the steepest descent algorithm. For each force field, the peptide was then relaxed by performing 50000 steps of MD simulation in the NVT ensemble and 50000 steps of simulation at 300 K and 1 bar in the NPT ensemble to equilibrate the temperature and density of the system respectively. The final production run was carried out in each force field for 500 ns at 300 K in the NPT ensemble.

All bonds were constrained using the LINCS algorithm[117] and a simulation time step of 2 fs was used. Periodic boundary conditions were applied in all directions. Lennard-Jones interactions were cut-off at 1.0 nm. Electrostatic interactions were

handled using the particle mesh Ewald approach with a real-space cut-off of 1.0 nm. The temperature was controlled using velocity rescaling with a time constant of 0.1 ps,[114] and the pressure was isotropically maintained at 1 bar using the Parrinello-Rahman barostat with a time constant of 2.0 ps and compressibility of 4.5×10^{-5} bar^{-1}.[113,115] Atomic coordinates were saved every 10 ps.

Ramachandran plots were used to compare relative helicity throughout the simulation. Based on this data, CHARMM27 was selected for subsequent simulations.

Following equilibration as previously described, the parent peptide was simulated for a total of 2 μs in the *NPT* ensemble, which was not sufficient for helical unwinding to occur, thus brute force MD simulations were deemed to be not computationally feasible, as it would be unlikely to reflect the experimentally observed differences in helicity between the unmodified and oxetane-modified peptides.

3.4.5 Steered MD of Helical Peptides

3.4.5.1 Model Building

Starting configurations of the peptides were generated using the Avogadro program version 1.2.0.[139] The peptides were built assuming an ideal α-helical conformation (ϕ and ψ backbone dihedral angles were set to -60° and -40° respectively) and uncharged ends. The ends were then replaced with N-terminal acetyl and C-terminal amide caps. For the oxetane-modified peptides, the carbonyl of either residue 3 or 8 was replaced by an oxetane ring. This was followed by a brief steepest descent energy minimisation in Avogadro using the Universal Force Field,[144] which generated the starting structure of each helical peptide for subsequent MD simulation in GROMACS 5.1.4.[137] Input topology files for GROMACS were generated using the GROMACS pdb2gmx tool.

3.4.5.2 Simulation Parameters

Prior to solvation, each structure was relaxed by performing energy minimisation. The structures were solvated with a pre-equilibrated box of 4931–4935 MeOH molecules and 4 chloride ions were added to balance the charge. The initial structures were relaxed by performing 50000 steps of steepest descent energy minimisation in vacuum. The final structures were then solvated with MeOH and the system was subjected to a further 50000 steps of steepest descent energy minimization. The systems were equilibrated in two stages, with position restraints of 1000 kJ mol^{-1} applied to peptide atoms. This equilibration consisted of 50000 steps of simulation at 310 K in the *NVT* ensemble, followed by 50000 steps of sim-

ulation at 310 K and 1 bar in the *NPT* ensemble to equilibrate the temperature and density of the system respectively.

Production MD simulations were performed for 40 ns and position restraints were released except for the Cα atom of Lys1, which was used as an immobile reference for the pulling simulations. For each helical structure, the Cα atom of Tyr18 was pulled along the z-axis, using a spring constant of 25 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and a pull speed of 0.00025 $\mathrm{nm\,ps^{-1}}$. These parameters were selected following a series of tests in which spring constants of 100 to 10 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and pull speeds of 0.02 to 0.00005 $\mathrm{nm\,ps^{-1}}$ were assessed using the parent peptide, the results of which are described in Chapter 5.

In all steered MD simulations, all bonds were constrained using the LINCS algorithm[117] and a simulation time step of 2 fs was used. Periodic boundary conditions were applied in all directions. Lennard-Jones interactions were cut-off at 1.0 nm. Electrostatic interactions were handled using the PME approach with a real-space cut-off of 1.0 nm. In initial temperature equilibration simulations, the temperature was controlled using a modified Berendsen thermostat with a time constant of 0.1 ps.[114] For the density equilibrations, temperature was maintained using the Berendsen thermostat and the pressure was isotropically maintained at 1 bar using the Berendsen barostat with a time constant of 2.0 ps and compressibility of 4.5×10^{-5} $\mathrm{bar^{-1}}$.[112] In the steered MD production run, temperature was controlled using the Nosé-Hoover thermostat.[113]

3.5 Microbiological Assays

All reagents were obtained from Sigma-Alrich, UK, unless otherwise specified.

3.5.1 Minimum Inhibitory Concentration

The minimum inhibitory concentration (MIC) is the lowest concentration of a chemical that prevents the visible growth of bacteria over a defined incubation period. The MIC is determined using dilution methods, by exposing the organism to twofold dilutions of a compound, and the result depends on the compound used, the strain of bacteria screened against (different strains may be resistant to different chemicals) and the composition of the growth medium, amongst other factors.[149] The lower the MIC, the more potent the compound is at inhibiting bacterial growth.

Initially we selected bacaucin-1a[150] (ALLSRVD) as a model system, and screened a variety of conditions and strains for this compound (Table 3.5). However, after being unable to observe any activity for this compound or bacaucin-1[151] (ELL-

SRVD), five new model systems were selected (Table 3.6). These were screened using the conditions listed in Table 3.7. MICs were observed for three model systems (Ac-RRWWCA-NH_2, GLLKRIK-NH_2 and KNKGWWW) against *E. coli* ATCC 25922 in M9 minimal media in the presence of DMSO, so these conditions were selected moving forward.

Table 3.5: Conditions and strains screened against for the bacaucin-1 and bacaucin-1a peptides.

Strain	Media	Additional conditions
E. coli ΔTolC	Cation-adjusted Mueller-Hinton broth	Peptides dissolved in DMSO
E. coli ATCC 25922	M9 minimal media (+ maltose)	0.1% TWEEN
S. aureus ATCC 29213	$\frac{1}{2}$ strength peptone water (+ 2% lactose)	

Table 3.6: Subsequent panel of five AMPs, with notes on published activity and references. Abbreviations in brackets are used throughout Chapter 6.

Peptide sequence	Activity against	Reference
Ac-RRWWCA-NH_2 (RRW)	*S. aureus* *E. coli* *S. sanguinis* *P. aeruginosa* *C. albicans*	Blondelle et al.[152]
GLLKRIK-NH_2 (GLL)	*S. aureus* ATCC 25923 *E. coli* ATCC 25922	Ifrah et al.[153]
KNKGWWW (KNK)	*S. aureus* ATCC 29213 *E. coli* ATCC 25922	Pasupuleti et al.[154]
ALLRL (ALL)	Gram+ bacteria	Otsuka et al.[155]
KVFLGLK (KVF)	*S. aureus* ATCC 25923 *S. typhimurium* *P. aeruginosa* *S. dysenteriae* *B. subtilis* *S. pneumoniae*	Xiao et al.[156]

Table 3.7: Conditions and strains screened against for the subsequent panel of five AMPs.[152–156]

Strain	Media	Additional conditions
E. coli ΔTolC	Cation-adjusted Mueller-Hinton broth	2–5% DMSO
E. coli ATCC 25922	M9 minimal media (+ maltose)	0.25–1 mM DTT
S. aureus ATCC 29213	M9 (+ GC media)	

MICs were determined using a broth microdilution method[157] in M9 minimal media supplemented with 5 g of maltose per litre of media. Prior to use, the media was filter-sterilised using a 0.22 μm filter. Peptides were dissolved in sterile distilled water in low-binding polypropylene Eppendorfs. Peptide solutions and quality control antibiotic (carbenicillin; Fisher Scientific, UK) dissolved in appropriate solvent[158] (water) were added to sterile flat-bottom polystyrene 96-well microplates at concentrations of approximately 0.5–256 μg/ml. Dithiothreitol (DTT) was added to each well containing the peptide Ac-RRWWCA-NH_2 to a final concentration of 1 mM, to prevent disulphide bond formation. This concentration of DTT was not sufficient to prevent bacterial growth. The final concentration of bacterial inoculum in each well was 7.5 $\times 10^5$ CFU/ml. DMSO was added to each well, corresponding to 2% of the final well volume. MIC was defined as the lowest concentration of peptide or antibiotic which inhibited the visible growth of bacteria after 18–20 hours at 37 °C. A separate biological repeat was carried out to confirm activity.

Following confirmation that RRW demonstrated antimicrobial activity, five additional sequences were screened (cRRWWCA, Ac-RRWWGA-NH_2, cRRWWGA, cRRWWG_{az}A and RWWG_{az}AR) to assess how cyclisation and azetidine modification alters antimicrobial activity. We used the same conditions as previously described, but omitted DTT, except where the peptide sequence contained a cysteine.

3.5.2 Minimum Bactericidal Concentration

The minimum bactericidal concentration (MBC) is the lowest concentration of a chemical required to kill a particular bacterium, and can be determined by subculturing broth dilutions taken at or above the MIC. The MBC is the lowest broth dilution of antimicrobial that prevents growth of the bacteria on the agar plate, as failure to grow on an agar plate that does not contain the test agent implies that only nonviable organisms are present.[149]

MBC assays were performed by transferring broth dropwise from the microplate following MIC assay onto an agar plate. The plates were incubated for 24 hours at 37 °C, and the MBC was defined as the lowest concentration for which there was no visible growth.

3.5.3 Toxicity Studies

In order for an AMP to be an effective therapeutic, it is important that it selectively targets prokaryotic cells and does not damage eukaryotic cells. One preliminary method to determine toxicity against mammalian cells is a haemolysis assay, in

which the compounds are incubated with red blood cells. Lysis at low concentrations or at concentrations close to the MIC indicates that the compound will not be suitable for therapeutic use.

Haemolytic activity was determined for RRW derivatives following a previously published procedure.[159] Defibrinated horse blood (E&O Laboratories Limited, UK) was centrifuged at 1,000 × g for 10 minutes and the supernatant and the buffy coat (the fraction that contains white blood cells and platelets) were removed. The erythrocytes were washed three times with cold 0.9% saline solution and resuspended to a concentration of 5% in 0.9% saline. Peptide solutions were added to sterile round-bottom polystyrene 96-well microplates at concentrations of approximately 2–1024 μg/ml. Controls of 0.9% saline and 5% Triton X-100 were used for 0% and 100% haemolysis respectively. Erythrocytes were then added to each well and incubated at 37 °C for one hour. Lack of haemolysis was noted by formation of a 'button' of erythrocytes in the bottom of the well.

Chapter 4

Heterocycle Modification on Small Peptides

4.1 Introduction

Peptide-based drugs remain promising as drug molecules, due to their potential to combine the advantages of traditional small molecule drugs with those of larger protein-based biologics.[16] For example, they have high potency and selectivity, have potentially lower toxicity than small molecule drugs, and can be readily synthesised. However, there are significant challenges surrounding the use of peptide-based drugs—they often have poor oral bioavailability, poor membrane permeability and are not resistant to proteases.[16,21,26] Modification of the peptide backbone via macrocyclisation or introduction of non-natural backbone linkages such as thioamides, azapeptides, and poly-N-substituted glycines has led to the increased stability or enhanced bioavailability of a wide range of peptides.[2–6]

The oxetane modification has previously been used in medicinal chemistry, often as bioisosteres for carbonyl or *gem*-dimethyl functional groups (Fig. 4.1).[53] Oxetanes also have emerging applications in peptide science,[63,160,161] and recently both Shipman[7] and Carreira[62] independently developed oxetane-modified peptidomimetics, where the carbonyl of a peptide backbone is replaced with a 3-aminooxetane unit. Oxetane-modified peptides could potentially have enhanced physicochemical properties and perform better as peptide-based drugs. For example, the serum half-life and *in vivo* analgesic properties of Leu-enkephalin, a short linear peptide neurotransmitter, can be improved by oxetane modification.[8,63]

Shipman and co-workers showed that oxetane-modified tripeptides with the sequence Leu-Gly-Ile (LGI) displayed antiparallel sheet-like arrangements in the solid state, and molecular dynamic (MD) simulations indicate that this oxetane-modified tripeptide has greater conformational flexibility compared to the unmod-

ified parent tripeptide. The unmodified tripeptide adopts extended conformations in which the C- and N-termini are separated by > 7 Å, whilst oxetane-modified LGI adopts a folded conformation in which the C- and N-termini are within 3–4 Å. The ability of the oxetane-modified tripeptide to more readily adopt a turn-like feature likely arises from the change in hybridisation and dihedral angle at the central amino oxetane unit.[7]

These observations suggest that oxetane modification could give rise to turn-like conformations in longer peptidomimetics. This could help improve efficiency of macrocylisation, a process which remains challenging, especially for very small compounds (less than seven amino acids), except where the peptide contains secondary structural elements such as β-turns.[49,50] Compared to their linear counterparts, cyclic peptides benefit from enhanced cell permeability, increased target affinity, and resistance to degradation by proteases, which makes them desirable drug candidates.[162,163]

Although oxetane-modified peptidomimetics have potential advantages compared to their unmodified counterparts, they are not without problems. The oxetane is prone to ring-opening under acidic conditions,[53] which limits the scope of amino acids that oxetane-modified peptides can contain, as all of the amino acid sidechain protecting groups used are removable using acids.[64] For Tyr(tBu), which can be removed under low acid concentrations, it is possible to achieve a reasonable yield of oxetane-modified cyclic peptides using 15% trifluoroacetic acid (TFA). However, for protecting groups such as Thr(tBu), which requires a higher concentration of TFA to remove, the yield decreases significantly,[41] and Arg(Pbf) is particularly challenging, as it requires more than 70% TFA to remove, resulting in very poor yields of oxetane-modified peptides.[92] For this reason, recently Shipman and co-workers have been investigating the impact of azetidine modification on peptide backbones.

Azetidines (Fig. 4.1), like oxetanes, have previously been used in medicinal chemistry.[66] They can enhance metabolic stability and pharmacokinetic properties,[67–69] and are found in several drugs (e.g. Baricitinib, Cobimetinib)[70,71] and biologically active natural products (e.g. Gelsemoxonine).[72] Azetidines have recently attracted significant interest due to their ability to be functionalised.[73,74] By selective functionalisation of the azetidine nitrogen, it may be possible to prepare functional derivatives with a variety of applications.[75–77]

In this chapter, the impact of oxetane and azetidine modification is explored on a pentapeptide, Leu-Ala-Gly-Ala-Tyr (LAGAY), derived from medicinal herbs and previously used to study macrocyclisation.[164] This work is published in part in Roesner et al.[92] We used biophysical methods including circular dichroism (CD) and two-dimensional nuclear magnetic resonance (NMR) techniques to charac-

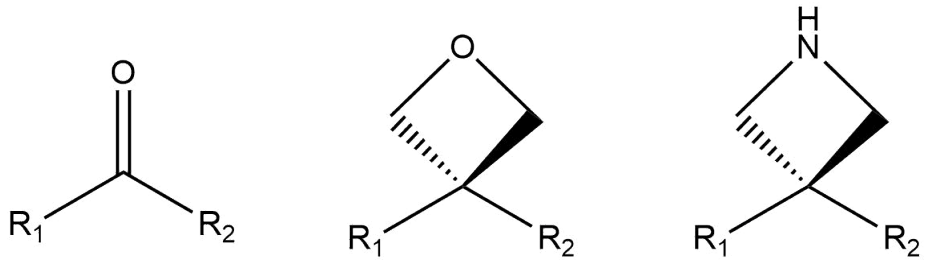


Figure 4.1: Comparison of carbonyl group, oxetane ring and azetidine ring.

terise differences between oxetane-modified and unmodified linear LAGAY, and inputted NMR-derived distance restraints into MD simulations. We found that oxetane modification on short linear peptides behaves in a turn-inducing manner, thereby increasing the efficiency of macrocyclisation.[92] We then characterised the effect of oxetane modification on cyclic LAGAY using CD, NMR and NMR-restrained MD. The MD simulations show clear differences in the structures of the modified and unmodified cyclic peptides. Finally, we also explored the effect of azetidine modification on the same linear peptide LAGAY. Using NMR techniques, we assessed how azetidine modification affects the structure of LAGAY, and found that it behaved in a similar manner to oxetane modification in this sequence.

4.2 Results and Discussion

4.2.1 Structural Impact of Oxetane Modification on Linear Precursors

Previous work by Shipman and co-workers suggests that oxetane-modified tripeptides adopt turn-like features.[7] We aimed to characterise the effect of oxetane on a longer, well-characterised sequence, LAGAY, and its oxetane-modified counterpart $LAG_{ox}AY$ (Fig. 4.2).

4.2.1.1 Circular Dichroism Reveals Changes in the Global Structure Upon Oxetane Modification

Circular dichroism (CD) spectroscopy was used to detect any changes in the global structure of linear peptides upon introduction of the oxetane. CD spectra were obtained for the linear precursor peptides $LAG_{ox}AY$ and $AG_{ox}AYL$ (Fig. 4.3). $AG_{ox}AYL$ contains the same amino acids as $LAG_{ox}AY$, but in a different order, so was used to assess whether the position of the oxetane residue changes the structure of the peptide. Spectra for unmodified LAGAY and AGAYL were not obtained due to technical difficulties (LAGAY sample contained residual moisture;

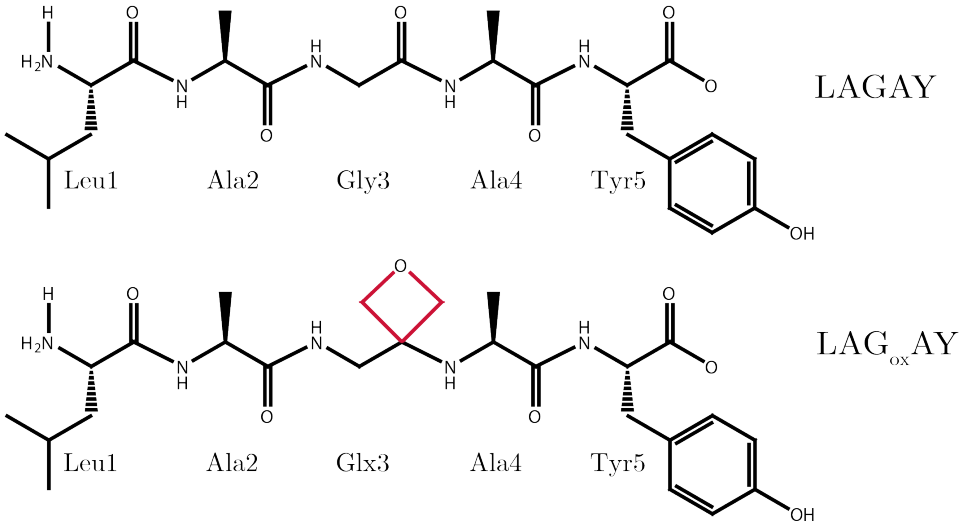


Figure 4.2: Structures of peptides LAGAY and LAG$_{ox}$AY. The oxetane is highlighted in red.

AGAYL contained 2 equivalents of DMF which strongly absorbs light < 240 nm).

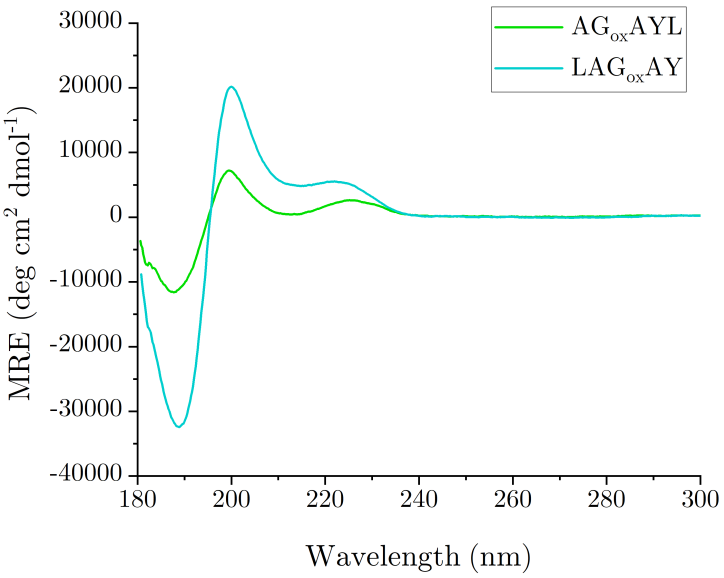


Figure 4.3: CD spectra for 175 μM oxetane-modified peptides AG$_{ox}$AYL (green) and LAG$_{ox}$AY (cyan) in 10 mM potassium phosphate, pH 7.0 at 25 °C.

Although it was not possible to obtain spectra for LAGAY and AGAYL, data for these peptides has previously been published by Tang et al., and LAGAY was shown to have a maximum absorbency at 202 nm, with a shoulder at 225 nm,[164] similar to the spectra we obtained for LAG$_{ox}$AY and AG$_{ox}$AYL (Fig. 4.3). However, Tang et al. obtained spectra using methanol as a solvent, which causes changes in hydrogen bonding and secondary structure of peptides,[165] while we used 10 mM potassium phosphate buffer, so it is difficult to compare between the two data sets. AGAYL has two negative maxima at 202 and 218 nm, suggesting some degree of α-helicity.[164] Both linear modified peptides, LAG$_{ox}$AY and AG$_{ox}$AYL,

yielded very similar spectra (Fig. 4.3) with a maximum absorbency at 202 nm, indicative of a β-turn (II) structure. This supports previous results by Powell et al. that the oxetane modification encourages turn-formation in short peptides.[7] As it was not possible to calculate the concentration of the peptides from absorbance, the concentration was estimated from dry weight and thus some deviation in signal intensity is observed.

4.2.1.2 NMR Provides Evidence for the Formation of a Turn in Oxetane-Modified Peptides

During the cyclisation process, peptides are coupled to 1,2,3-Benzotriazin-4(3H)-one, derived from the reagent DEPBT. However, this 'activated ester' form of the peptide (Fig. 4.4) is highly reactive and cannot be isolated for NMR analysis. Thus, two-dimensional NMR experiments (see section 3.2.3) were performed on the linear, zwitterionic forms of the peptides to better understand the conditions prior to cyclisation.

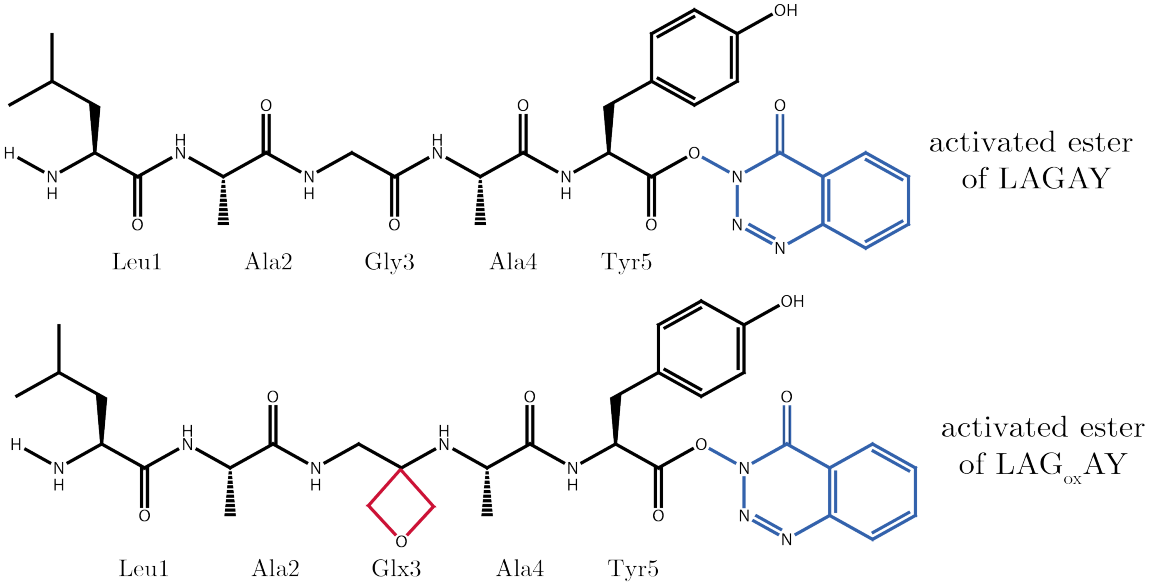


Figure 4.4: Structures of activated ester of LAGAY and $LAG_{ox}AY$. The 1,2,3-Benzotriazin-4(3H)-one is highlighted in blue, and the oxetane modification is highlighted in red.

2D ^{1}H-^{1}H TOCSY and NOESY spectra were acquired in order to assign the proton signals in both peptides and probe changes in conformation, especially near the site of modification. A larger number of nuclear Overhauser effects (NOEs) in oxetane-modified $LAG_{ox}AY$ were observed compared to the unmodified parent. While sequential NOEs are observable in both peptides (dαN, dNN, dβN), medium-range NOEs, commonly observed in α-helices, β-sheets, and turns,[132] are only observed in peptide $LAG_{ox}AY$ (Fig. 4.5).

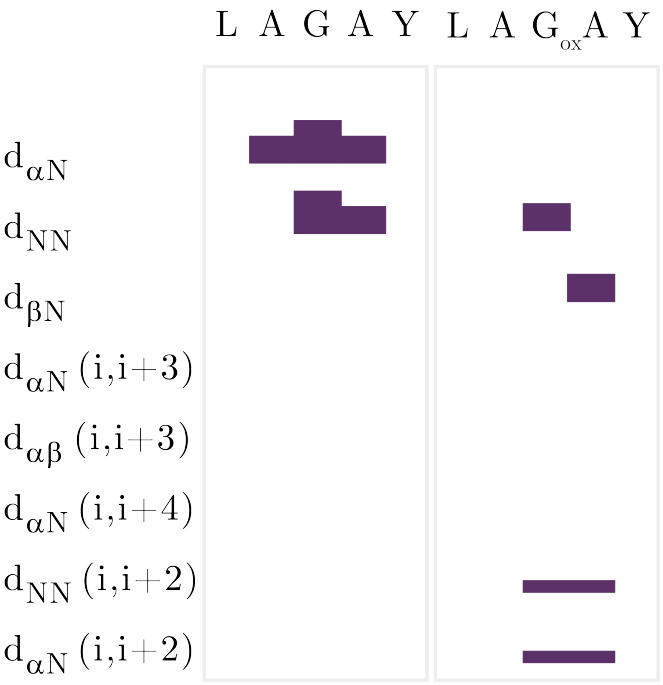


Figure 4.5: NOE-connectivity maps of LAGAY and LAG$_{ox}$AY based on NOE spectra collected with 250–800 ms mixing times. Each sequential and medium-range coupling is indicated on the left, and a bar is used to join the observed coupled residues.

To ensure that the previously observed turn-inducing effect is not due to charge interactions between the zwitterionic ends, linear LAGAY with a C-terminal methyl ester (Fig. 4.6a) was prepared. This methyl ester, LAGAY-OMe, and modified LAG$_{ox}$AY-OMe (Fig. 4.6b), should better mimic the behaviour of the activated ester, and prevent any charge-charge interactions between the termini.

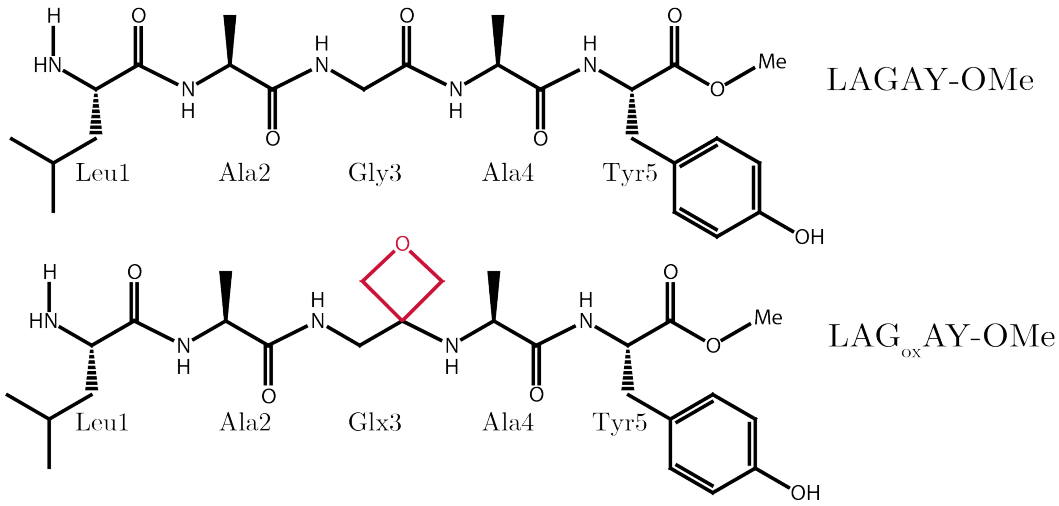


Figure 4.6: Structures of methyl esters LAGAY-OMe and LAG$_{ox}$AY-OMe. The oxetane is highlighted in red.

Like its zwitterionic counterpart, oxetane-modified LAG$_{ox}$AY-OMe yielded a larger number of NOEs than the unmodified parent (Fig. 4.7a and b), which were used to construct an NOE-connectivity map (Fig. 4.7c) summarising sequential and medium-range NOEs. Again, sequential NOEs are observable in both peptides,

and medium-range NOEs (dNN (*i, i*+2) and dαN (*i, i*+2)) are only observed in modified peptide $LAG_{ox}AY$-OMe. These longer-range NOEs are clearly observed in the oxetane-modified peptide between residues Gly3 and Tyr5, indicating their close proximity, suggesting the formation of a turn. These NMR studies provide the first direct experimental evidence that the oxetane modification is turn-inducing, and brings the peptide termini closer in space to facilitate efficient cyclisation.

Inter-residue NOE restraints were incorporated into MD simulations (Table 4.1). The GROMACS implementation follows that of Torda et al., whereby time-averaged distance restraints are used.[137,146] As an NOE may reflect an averaging of multiple conformations, time-averaged distance restraints provide a better approximation of the physical nature of the NOE by enabling an atom to satisfy seemingly incompatible distance restraints on average by moving between multiple positions.[146] If the NOE upper bound of the time-averaged distance between two atoms is exceeded, a harmonic restoring force (the strength of which is controlled by the corresponding force constant for the restraint and is described below) pulls the atoms back towards each other. Based on the analysis of the NMR experiments described above, NOEs were binned based on their intensities into strong, medium and weak. The distance ranges and force constants were 1.8–2.7 Å and 2000 $kJ\,mol^{-1}\,nm^{-2}$, for strong restraints, 1.8–3.3 Å and 1500 $kJ\,mol^{-1}\,nm^{-2}$ for medium restraints and 1.8–5.0 Å and 1000 $kJ\,mol^{-1}\,nm^{-2}$ for weak restraints.

Table 4.1: Inter-residue ^1H-^1H NOEs observed for pentapeptides LAGAY-OMe and $LAG_{ox}AY$-OMe. Intensities correspond to ^1H-^1H distances of 1.8–2.7 Å (strong), 1.8–3.3 Å (medium), and 1.8–5.0 Å (weak). NMR restraints were implemented in MD simulations with the above distance ranges and force constants of 2000 $kJ\,mol^{-1}\,nm^{-2}$ (strong), 1500 $kJ\,mol^{-1}\,nm^{-2}$ (medium) and 1000 $kJ\,mol^{-1}\,nm^{-2}$ (weak).

Peptide	NOE	Intensity
LAGAY-OMe	Ala2Hα–Gly3NH	Weak
	Gly3Hα–Ala4NH	Weak
	Ala4Hα–Tyr5NH	Weak
$LAG_{ox}AY$-OMe	Ala2NH–Glx3Hα	Medium
	Leu1Hα–Ala2NH	Weak
	Ala2Hα–Glx3NH	Weak
	Ala2Hβ–Glx3NH	Weak
	Glx3NH–Tyr5NH	Weak
	Glx3Hα–Tyr5NH	Weak
	Ala4Hα–Tyr5NH	Weak

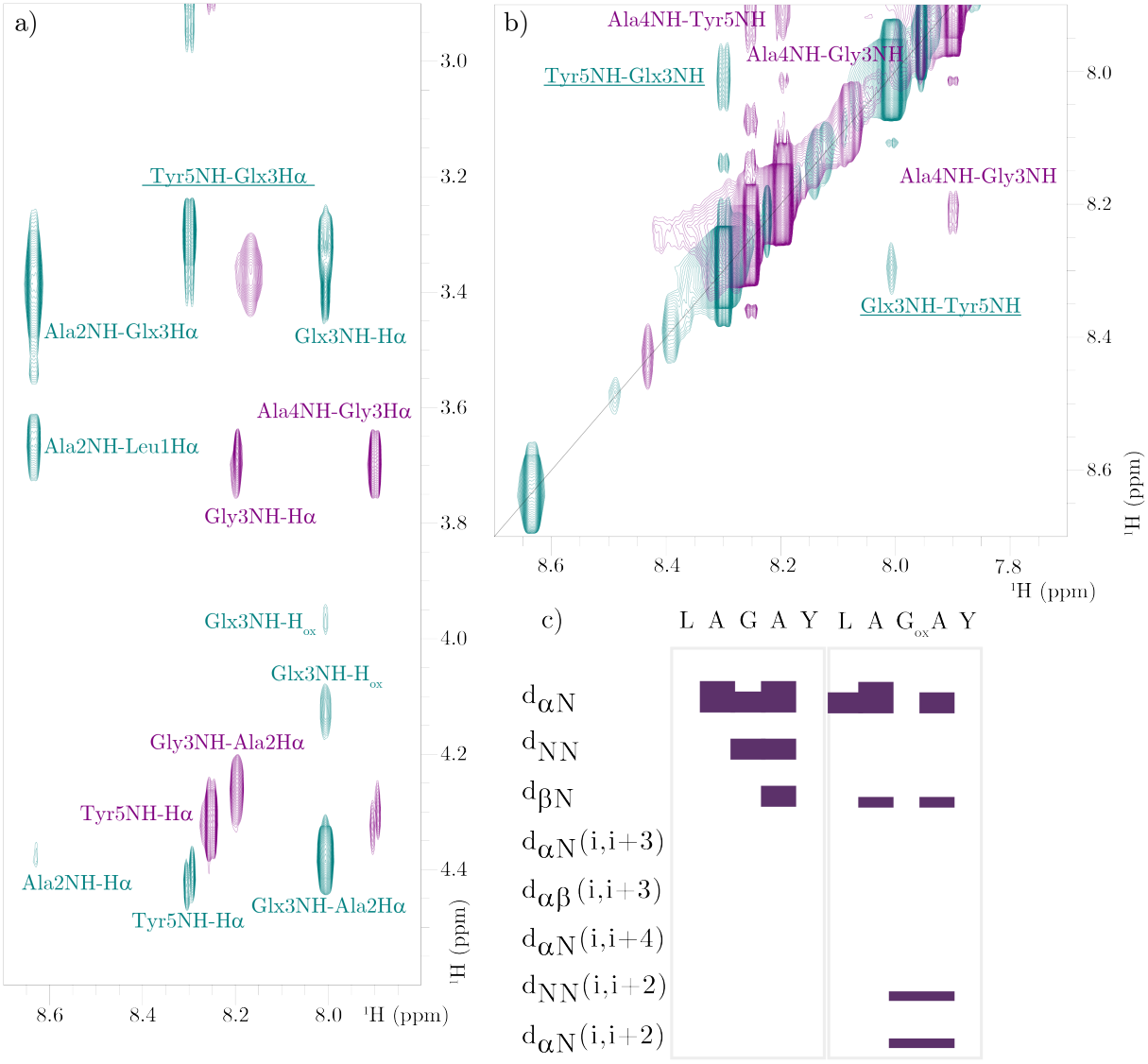


Figure 4.7: (a) Fingerprint region and (b) amide region of NOESY spectra of $LAG_{ox}AY$-OMe (cyan) and LAGAY-OMe (purple) collected with a mixing time of 400 ms for samples containing 30 mM peptide solubilised in d_6-DMSO. Both spectra were collected at 298 K at a field strength of 700 MHz. Medium-range NOEs are underlined. (c) NOE-connectivity maps of $LAG_{ox}AY$-OMe and LAGAY-OMe based on NOE spectra collected with 250–800 ms mixing times. Each sequential and medium-range coupling is indicated on the left, and a bar is used to join the observed coupled residues. Reproduced from Roesner et al.[92]

4.2.1.3 NMR-Restrained MD Simulations Show the Turn-Inducing Effect

The peptides were built in Avogadro 1.2.0[139] with a linear conformation and uncharged ends. The methyl ester end group was added to both peptides and C=O of peptide $LAG_{ox}AY$ was substituted with the oxetane group. Each peptide was briefly subjected to a steepest descent energy minimisation in Avogadro using the Universal Force Field.[144] This generated methyl ester peptide structures used in subsequent simulations in GROMACS 5.1.4.[137]

The initial structures were relaxed by performing 50000 steps of MD simulation in the *NVT* ensemble in vacuum at 300 K. The final structures were then solvated with DMSO (to better mimic the conditions used in the NMR experiments) and the system was subjected to 50000 steps of steepest descent energy minimization. This was followed by 100000 steps of simulation at 300 K in the *NVT* ensemble and 50000 steps of simulation at 300 K and 1 bar in the *NPT* ensemble to equilibrate the temperature and density of the system respectively. To overcome the issue of kinetic trapping in local minima and to enhance the sampling of conformational space, each peptide was then simulated for 100 ns at 500 K in the *NVT* ensemble. Cluster analysis was performed on the resultant trajectory to group peptide conformations according to their structural similarity, using an algorithm described by Daura et al.[147], and a cut-off of 0.15 nm. The central structure of the top five most populated clusters (which accounted for $> 99.9\%$ of the total population) was then used as the starting configuration for five independent 100 ns simulations of each peptide at 300 K and 1 bar in the *NPT* ensemble. Accordingly, each peptide was simulated for a total simulation time of 500 ns.

Cluster analysis was performed on 12500 structures extracted from the trajectories at 40 ps intervals, on sections of the trajectory from 0–20 ns initially, adding data in 20 ns at a time until the full 0–500 ns trajectory was analysed. By repeating the cluster analysis with different lengths of trajectory, it is possible to assess how the number (Fig. 4.8) and populations (Fig. 4.9) of clusters changes over time, and confirm that the data are converging.

$LAG_{ox}AY$-OMe appears to be much more flexible than the parent peptide—it takes significantly longer for the cluster populations to begin to converge (Fig. 4.9). Furthermore, although both peptides can be described as having four distinct structures, $LAG_{ox}AY$-OMe spends only 65.05% of frames in its most populated cluster, compared to 92.51% for LAGAY-OMe (Fig. 4.9 and 4.10). Intriguingly, $LAG_{ox}AY$-OMe spends $> 10\%$ of its time in a conformation where the termini are within 5 Å, whereas LAGAY-OMe spends a mere 0.02% of its time in a similar conformation (Fig. 4.10). This may provide an explanation as to why oxetane

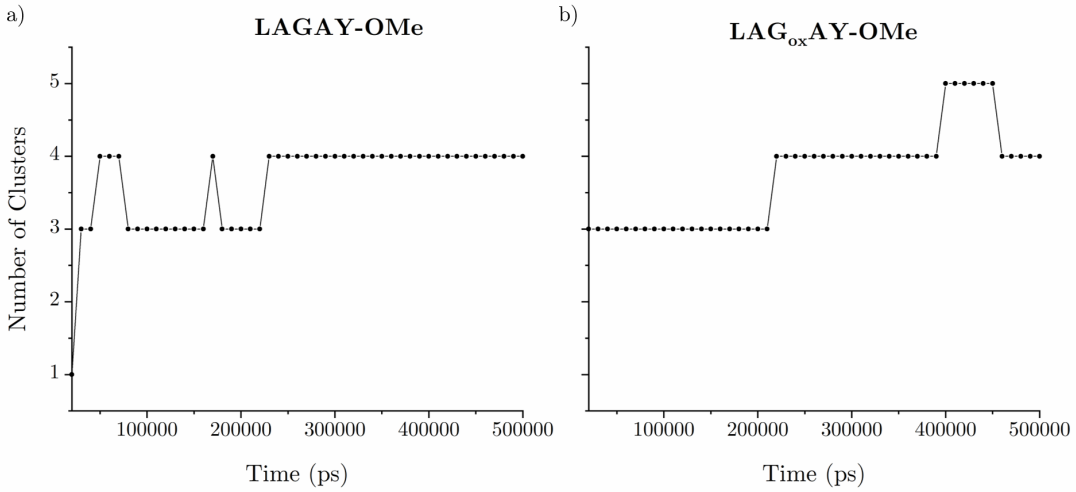


Figure 4.8: Number of clusters over time as the five independent 100 ns simulations are added to the trajectory for (a) the parent peptide LAGAY-OMe, and (b) the oxetane-modified peptide LAG$_{ox}$AY-OMe.

modification is able to so drastically improve cyclisation efficiency, and these MD data are in strong agreement with the NMR experiments which provide evidence of turn formation surrounding the oxetane.

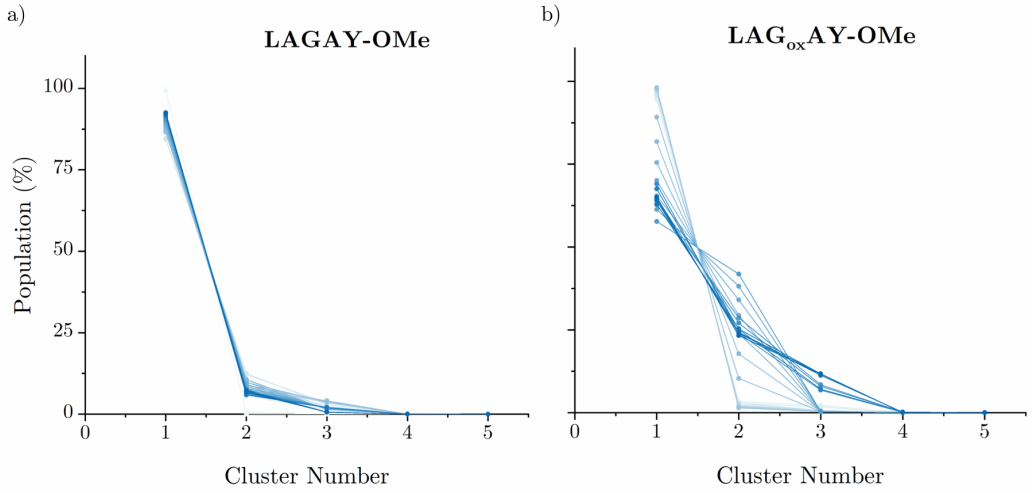


Figure 4.9: Population (% of total) of each cluster as the 5 independent 100 ns simulations are added to the trajectory for (a) the parent peptide LAGAY-OMe, and (b) the oxetane-modified peptide LAG$_{ox}$AY-OMe. Data are overlaid going from the lightest shade of blue 0–20 ns to the darkest shade of blue 0–500 ns.

Figure 4.10: Snapshots of each of the clusters for (a) LAGAY-OMe, and (b) $LAG_{\alpha x}AY$-OMe. Single structures are shown for each cluster for each peptide. The population percentage (rounded to 2 d.p.) of each cluster is indicated above. Yellow dashed lines indicate where the termini are within 5 Å.

Similarly, Ramachandran plots (Fig. 4.11), showing the distribution ϕ and ψ backbone dihedral angles of the entire peptide (averaged over the entire simulation trajectory), indicate that LAGAY-OMe spends most of its time in a single dominant conformation, with additional minor conformations that sample a range of ϕ/ψ space. In contrast, the plot for modified peptide $LAG_{ox}AY$-OMe shows more 'hotspots', representing multiple higher population clusters.

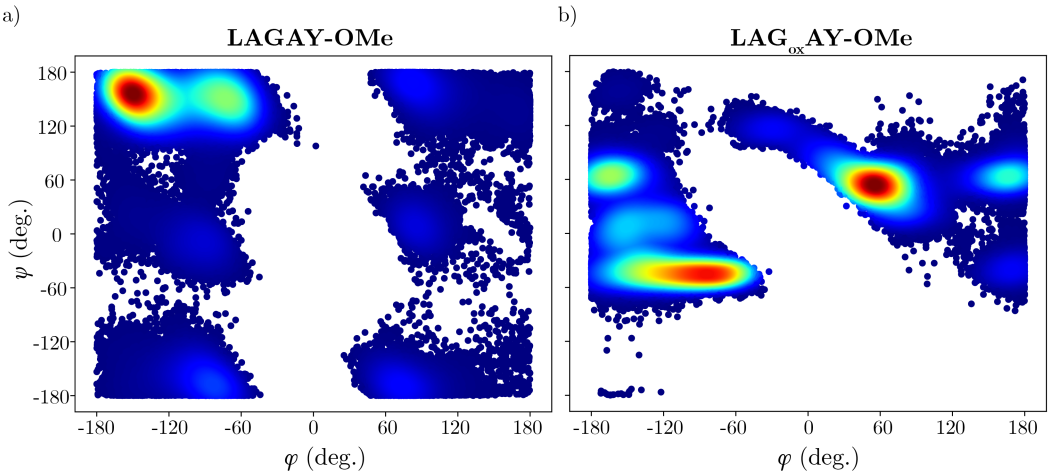


Figure 4.11: Ramachandran plots indicating differences in ϕ/ψ space for linear peptides LAGAY-OMe and $LAG_{ox}AY$-OMe.

The goodness-of-fit to the restraints in the resulting structures were evaluated (Table 4.2). There were no restraint violations for the unmodified peptide, but the oxetane-modified $LAG_{ox}AY$-OMe did experience some violations of the restraint between Ala2NH and Glx3Hα. However, these violations are within ranges similar to those recorded in previous MD simulations with NMR-derived restraints.[166] In future, the force constants selected for NMR-derived restrained may need further optimisation, as violations are only observed in the stronger restraint (and therefore only observed in the modified peptide, as the parent peptide only contains weaker restraints).

Table 4.2: Restraint violations (root mean squared deviation from experimental distance restraints (Å) with standard deviation) throughout simulations for LAGAY-OMe and $LAG_{ox}AY$-OMe.

Parameter	LAGAY-OMe	$LAG_{ox}AY$-OMe
All inter-residue	0.00 ± 0.01	0.14 ± 0.21
Medium	N/A	0.54 ± 0.11
Weak	0.00 ± 0.01	0.05 ± 0.06

Taken together, the data suggests that oxetane behaves in a manner that is consistent with a turn-inducing effect. From the MD simulations with NMR re-

straints, we see that for the modified peptide the termini spend a higher proportion of the time within close proximity. This work is in agreement with previously published simulation work involving oxetane-modified tripeptides,[7] and correlates well with other experimental observations that oxetane modification drastically increases efficiency of macrocyclisation.[92]

4.2.2 Structural Impact of Oxetane Modification on Cyclic LAGAY

Oxetane behaves as a turn-inducer in short linear peptides, and as a result it substantially increases the efficiency of macrocyclisation (by 20% for the sequence LAGAY, significantly more than other modifications assessed such as N-methylation).[92] We aimed to understand how oxetane modification may change the structure of the cyclic peptide, since a cyclic backbone is highly desirable when designing peptide-based drugs.[16] In order to investigate the effect of the oxetane modification on cyclic peptides, cyclic LAGAY (cLAGAY) and cyclic oxetane-modified LAGAY (cLAG$_{ox}$AY) were selected as model systems (Fig. 4.12), based on a sequence identified from medicinal herbs and previously used as a model system to explore influences on the factors of cyclisation.[164]

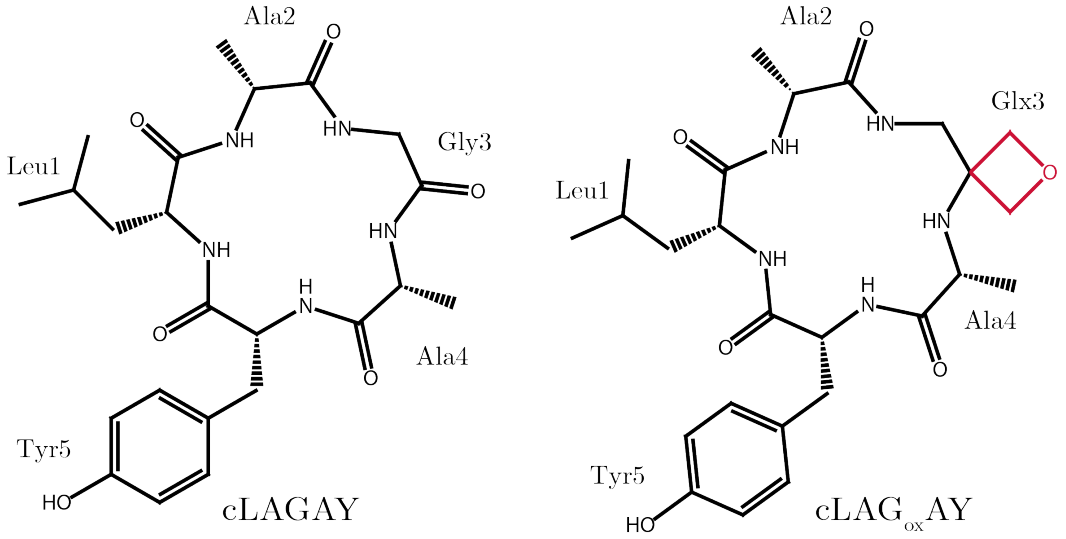


Figure 4.12: Structures of cLAGAY and cLAG$_{ox}$AY. The oxetane is highlighted in red.

4.2.2.1 Circular Dichroism Reveals Clear Differences Between cLAGAY and cLAG$_{ox}$AY

CD spectra were obtained for cyclic peptides cLAGAY and cLAG$_{ox}$AY (Fig. 4.13). For these cyclic peptides, there was no change in overall spectral shape at 10 °C

compared to 25 °C, although at lower temperatures the spectrum becomes noisier (data not shown). Therefore a temperature of 25 °C was selected.

Comparison of the spectra for cLAGAY and cLAG$_{ox}$AY (Fig. 4.13) show clear differences in structure upon introduction of the oxetane modification, likely reflecting changes in chirality for the cyclic peptides. The spectrum of cLAGAY shows negative peaks at 197 and 202 nm, and looks very similar to a previous spectrum published by Tang and colleagues, apart from the presence of a slight shoulder at ca. 218 nm, tailing off to 0 at 230 nm,[164] which was observed in their data but was not observed in our hands. This difference may be due to differences in solvent and concentrations—our spectra were obtained at a much lower concentration (175 μM vs. 4200–8400 μM), and in a different solvent (phosphate buffer vs. methanol).

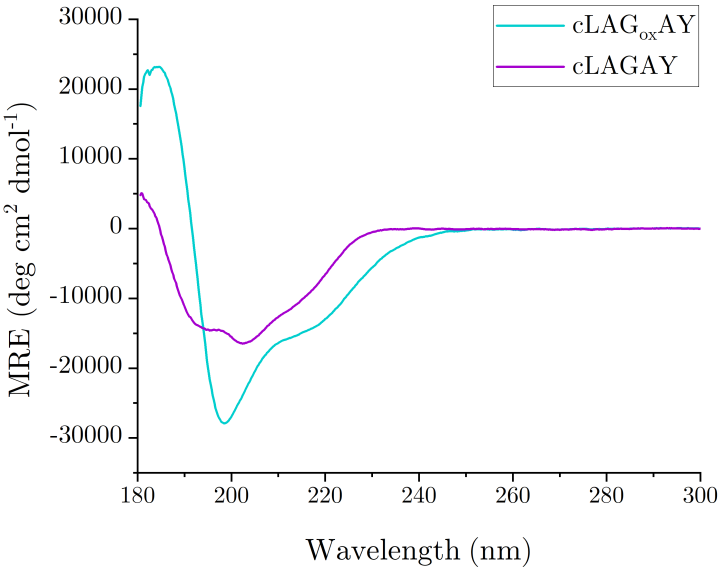


Figure 4.13: CD spectra for 175 μM cyclic peptides cLAGAY (purple) and cLAG$_{ox}$AY (cyan) in 10 mM potassium phosphate, pH 7.0 at 25 °C.

The spectrum for cLAG$_{ox}$AY is similar to the unmodified cyclic peptide spectra observed by Howell and colleagues.[167] Intriguingly, the spectra for AG$_{ox}$AYL and LAG$_{ox}$AY (Fig. 4.3) appear to be inverted compared to cLAG$_{ox}$AY, which may suggest opposite chirality,[168] but is more likely due to π-stacking of tyrosines occurring in the linear peptides, resulting in two overlapping excitons centred at 220 and 190 nm.

4.2.2.2 NMR Shows Oxetane Modification Alters the Structure of the Cyclic Peptide

In order to characterise the structural changes in these peptides with atomic resolution, two-dimensional ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY experiments were used to assign the ^{1}H atoms in cLAGAY and cLAG$_{ox}$AY and to compile a list of inter-

residue NOE-derived distance restraints. For both peptides full assignment was achieved (Tables 4.3 and 4.4).

Table 4.3: Chemical shifts of cyclic peptide cLAGAY in d_6-DMSO at 298 K. (*) indicates protons with degenerate chemical shifts.

Residue	Chemical shift (ppm)			
	NH	Hα	Hβ	others
Leu1	8.50	4.16	1.46	Hγ 1.63; Hδ 0.82, 0.88
Ala2	8.51	4.27	1.23	
Gly3	8.39	3.49, 3.85		
Ala4	8.15	4.10	1.14	
Tyr5	8.23	4.36	2.88	Hδ 7.00*; Hε 6.64*; Hη 9.22

Table 4.4: Chemical shifts of cyclic peptide cLAG$_{ox}$AY in d_6-DMSO at 298 K. (*) indicates protons with degenerate chemical shifts. H_{ox} refers to the four protons of the oxetane ring.

Residue	Chemical shift (ppm)			
	NH	Hα	Hβ	others
Leu1	8.49	4.16	1.49	Hγ 1.62; Hδ 0.80, 0.89
Ala2	8.51	4.30	1.29	
Glx3	7.34	3.42, 3.49		H_{ox} 4.00, 4.10, 4.19, 4.44
Ala4	3.30	2.72	1.04	
Tyr5	8.49	4.42	2.96	Hδ 7.00*; Hε 6.64*; Hη 9.19

Comparison of these assignments as well as representative NOESY overlays from the fingerprint region (Fig. 4.14a) and near the diagonal (Fig. 4.14b) reveal distinct changes in chemical shifts observed upon addition of an oxetane ring. For example, in cLAG$_{ox}$AY, Glx3 and Ala4 amide protons are shifted upfield due to shielding by the oxetane ring. The oxetane ring of the modified glycine has a distinct signature of four peaks around the diagonal.

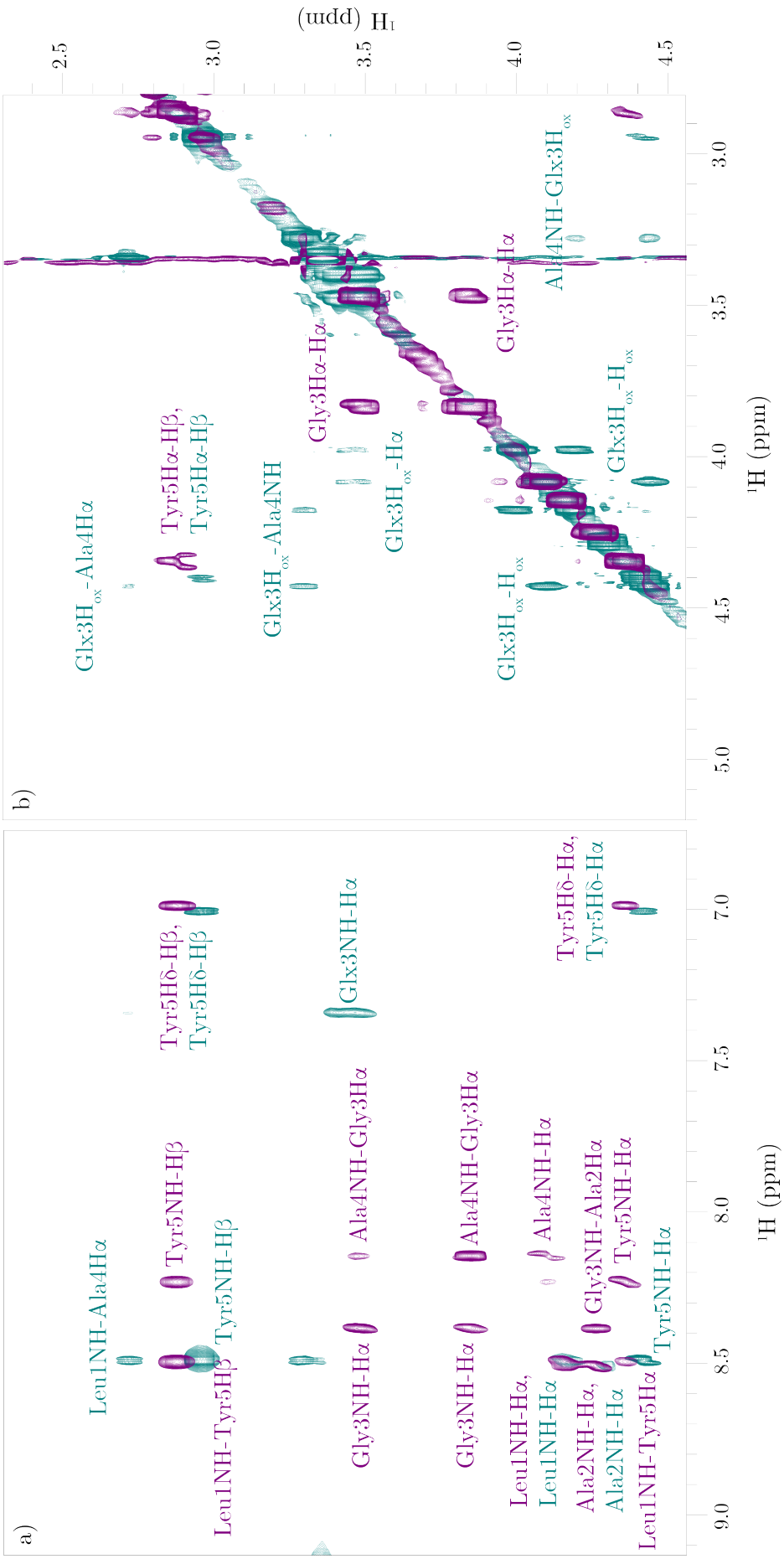


Figure 4.14: (a) Fingerprint region and (b) amide region of NOESY spectra of cLAG$_{ox}$AY (cyan) and cLAGAY (purple) collected with a mixing time of 250 ms for samples containing 60 mM peptide solubilised in d_6-DMSO. Both spectra were collected at 298 K at a field strength of 700 MHz.

To ensure that NOE restraints were estimated using a mixing time at which spin diffusion is not occurring, NOE build-up curves (NOE intensity vs. mixing time) were plotted for inter-residue NOE peaks following normalisation to the Tyr aryl peak volume (Fig. 4.15a). The Tyr aryl has a fixed ^{1}H-^{1}H distance and a well-characterised NOE peak which correlates with mixing time, so the aryl peaks can be used to normalise data in which peak intensity varies due to changes in receiver gain.

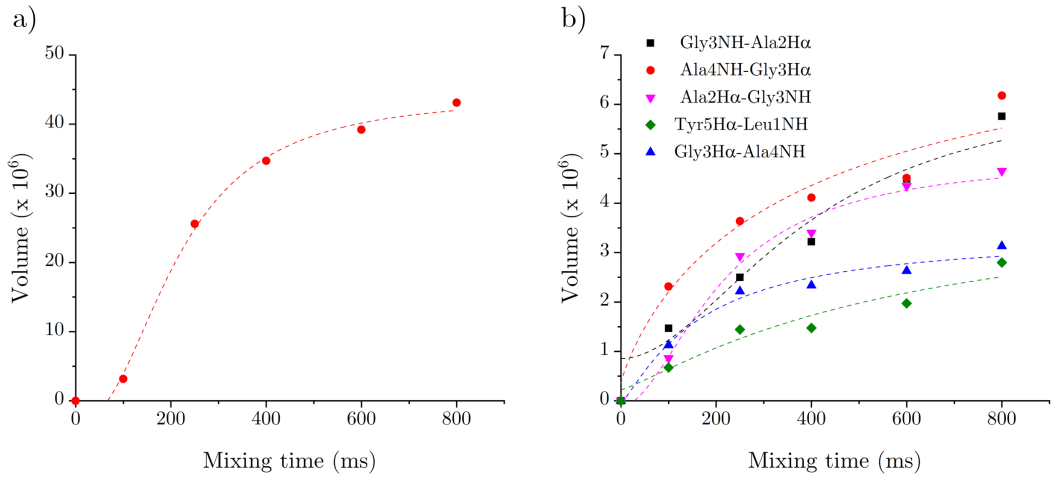


Figure 4.15: NOE build-up curves for cLAGAY (a) tyrosine aryl protons, and (b) inter-residue NOEs. Inter-residue NOE peak volumes were normalised against the Tyr aryl peak volume (due to differing peak intensities caused by changes in receiver gain of the spectrometer) and were plotted for mixing times 100–800 ms.

The NOE build-up curves shown in Fig. 4.15b suggest that a mixing time of 100–250 ms would be appropriate for selection of distance restraints, as in this region the peak volumes correlate linearly with mixing time, thus a mixing time of 250 ms was selected for estimation of NOE restraints. Inter-residue restraints were incorporated into MD simulations (Table 4.5).

A strong NOE peak corresponding to Ala4Hα-Tyr5OH was observed in the NOE spectra of cLAG$_{ox}$AY. Dipolar coupling of these protons was unexpected, and to ensure this NOE peak was not a result of an inter-molecular interaction between peptide molecules, a series of dilution experiments was performed. 1D ^{1}H NMR spectra were recorded for both cLAGAY and cLAG$_{ox}$AY at concentrations of 2–60 mM (Fig. 4.16 and 4.17). At 2 mM, TOCSY and NOESY experiments were also performed, although these spectra had poor signal-to-noise ratios (data not shown). There are slight changes in chemical shifts as concentration decreases corresponding to the amide protons of the peptide. However, the chemical shift of Tyr5OH is unchanged from concentrations of 2–60 mM (Fig. 4.17), suggesting there are no inter-molecular interactions occurring at this site.

Table 4.5: Inter-residue ^{1}H-^{1}H NOEs observed for cyclic pentapeptides cLAGAY and cLAG$_{ox}$AY. Intensities correspond to ^{1}H-^{1}H distances of 1.8–2.7 Å (Strong), 1.8–3.3 Å (Medium), and 1.8–5.0 Å (Weak).

Peptide	NOE	Intensity
cLAGAY	Leu1NH–Tyr5NH	Strong
	Leu1NH–Tyr5Hβ1	Medium
	Leu1NH–Tyr5Hβ2	Medium
	Gly3NH–Tyr5NH	Medium
	Leu1NH–Tyr5Hα	Weak
	Gly3NH–Ala2Hα	Weak
	Gly3Hα1–Ala4NH	Weak
	Gly3Hα2–Ala4NH	Weak
cLAG$_{ox}$AY	Ala4Hα–Tyr5OH	Strong
	Glx3NH–Tyr5NH	Medium
	Glx3NH–Ala4NH	Weak
	Glx3H$_{ox}$–Ala4Hα	Weak
	Glx3H$_{ox}$–Ala4NH	Weak
	Ala4Hα–Leu1NH	Weak

4.2.2.3 MD Simulations with NMR-Derived Restraints Show Changes in Hydrogen Bonding Upon Oxetane Modification

The peptides were built in Avogadro 1.2.0[139] with a linear conformation and uncharged ends. Both peptides were then cyclised and C=O of peptide cLAG$_{ox}$AY was substituted with the oxetane group. Each peptide was briefly subjected to a steepest descent energy minimisation in Avogadro using the Universal Force Field.[144] This generated cyclic peptide structures used in subsequent simulations in GROMACS 5.1.4.[137]

The simulations were performed as for the linear peptides described in section 4.2.1.3, except the cut-off for the cluster analysis was 0.05 nm, due to the relative rigidity of the cyclic backbone compared to the linear counterparts. The central structure of the top five most populated clusters (which accounted for $> 99.9\%$ of the total population) was then used as the starting configuration for five independent 100 ns simulations of each peptide at 300 K and 1 bar in the *NPT* ensemble. Accordingly, each peptide was simulated for a total simulation time of 500 ns.

As for the linear peptides, cluster analysis was carried out on sections of the trajectory. The numbers (Fig. 4.18) and population distributions (Fig. 4.19) of the different clusters throughout the length of the trajectory indicate simulation convergence.

As a smaller cut-off was used during the cluster analysis due to the comparative rigidity of the cyclic peptides, it is not possible to compare the cluster analysis of the cyclic peptides to their linear precursor counterparts. However, we can com-

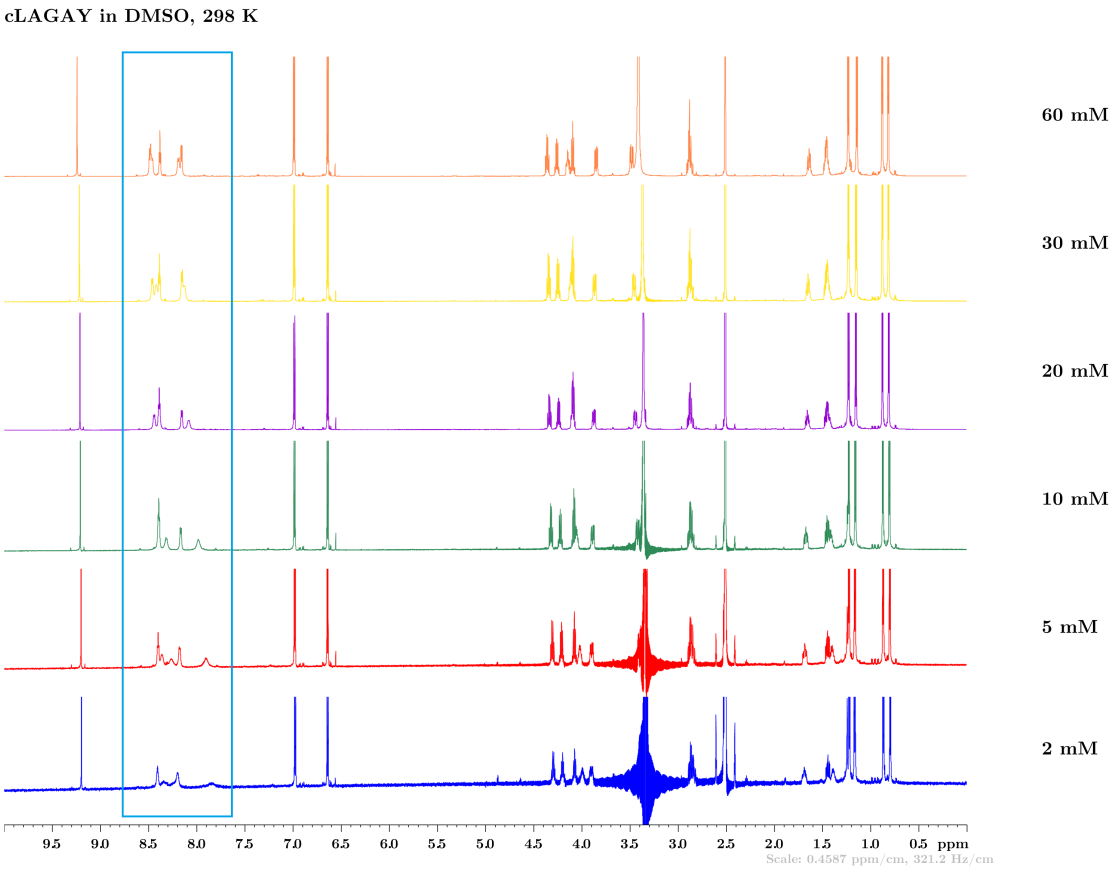


Figure 4.16: 1D ^{1}H NMR experiments of cLAGAY at concentrations of 2–60 mM. Changes appear to be occurring in the amide region of the spectra (indicated with blue box) and should be further investigated.

pare unmodified cLAGAY to oxetane-modified peptide cLAG$_{ox}$AY. Five and four distinct structures were found for peptides cLAGAY and cLAG$_{ox}$AY respectively, with the first three clusters representing $\geq$ 99.9% of the entire trajectory for both peptides. The trajectories were dominated by conformations belonging to a single cluster that represented 93.7% and 99.1% of the total population of cLAGAY and cLAG$_{ox}$AY, respectively. The cluster analysis indicates that the backbones of both structures have little conformational flexibility (as expected for small cyclic peptides), but that cLAGAY is more flexible than the oxetane-modified cLAG$_{ox}$AY, as cLAGAY samples a greater number of clusters (Fig. 4.18a and b), and the population of the most dominant structure is reduced compared to cLAG$_{ox}$AY (Fig. 4.19a and b). In contrast, cLAG$_{ox}$AY is rigid, and spends over 99% of the total trajectory in a single backbone conformation (Fig. 4.20).

Similarly, Ramachandran plots (Fig. 4.21), showing the distribution ϕ and ψ backbone dihedral angles of each residue and the entire peptide (averaged over the entire simulation trajectory), indicate that cLAGAY samples more ϕ/ψ space than cLAG$_{ox}$AY, suggesting decreased flexibility on addition of oxetane. However, this

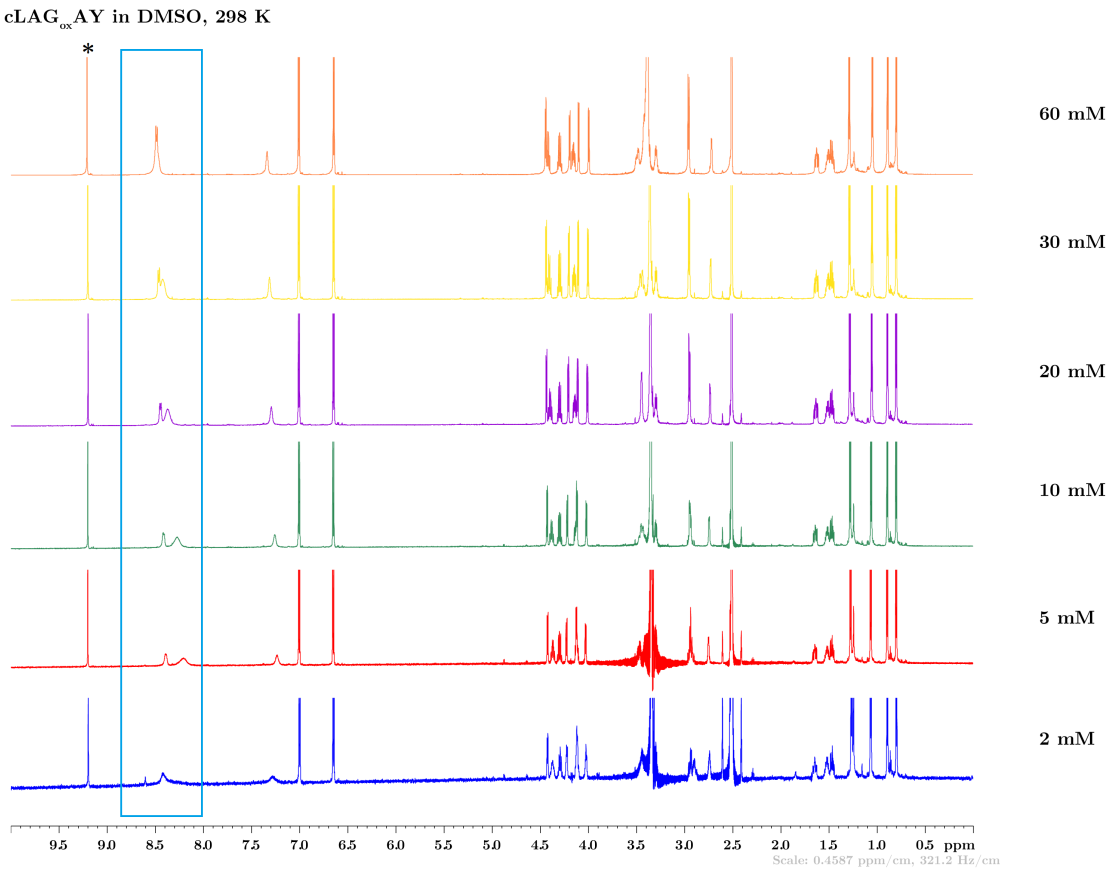


Figure 4.17: 1D ^{1}H NMR experiments of cLAG$_{ox}$AY at concentrations of 2–60 mM. Changes appear to be occurring in the amide region of the spectra (indicated with blue box) and should be further investigated. The peak corresponding to Tyr5OH is indicated with (*).

is in contrast to what has been observed previously, where oxetane modification resulted in increased backbone flexibility of oxetane-modified peptides, as this module acts as a β-turn-inducing element.[7]

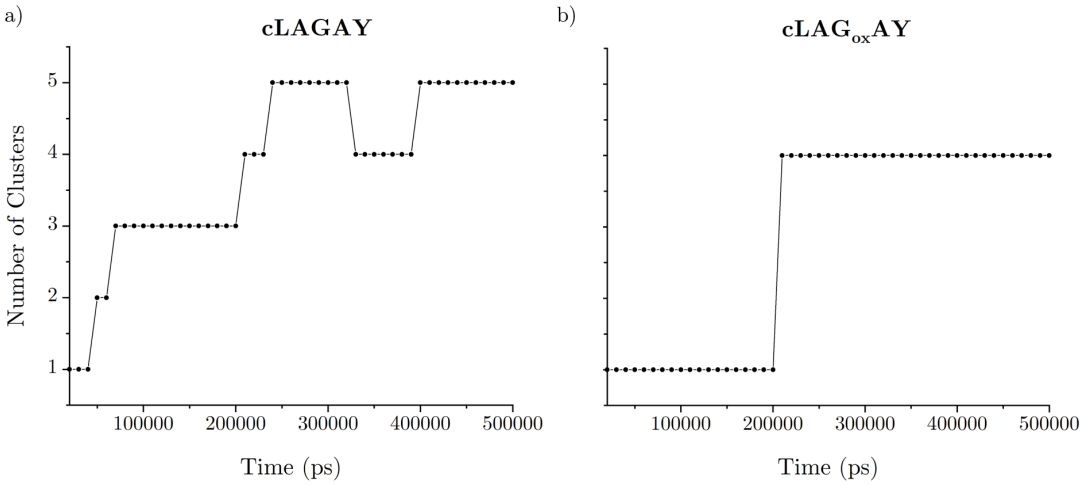


Figure 4.18: Number of clusters over time as the five independent 100 ns simulations are added to the trajectory for (a) the parent peptide cLAGAY, and (b) the oxetane-modified peptide cLAG$_{ox}$AY.

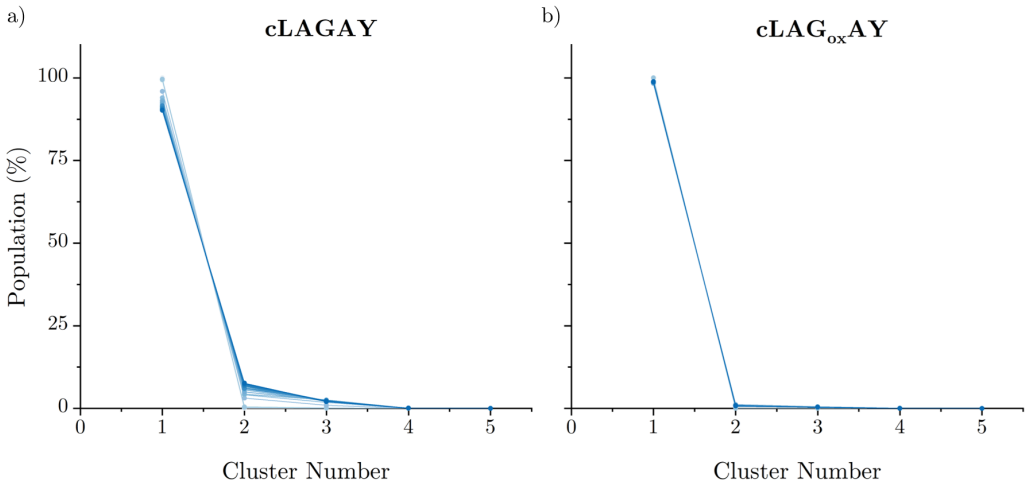


Figure 4.19: Population (% of total) of each cluster as the 5 independent 100 ns simulations are added to the trajectory for (a) the parent peptide cLAGAY, and (b) the oxetane-modified peptide cLAG$_{ox}$AY. Data are overlaid going from the lightest shade of blue 0–20 ns to the darkest shade of blue 0–500 ns.

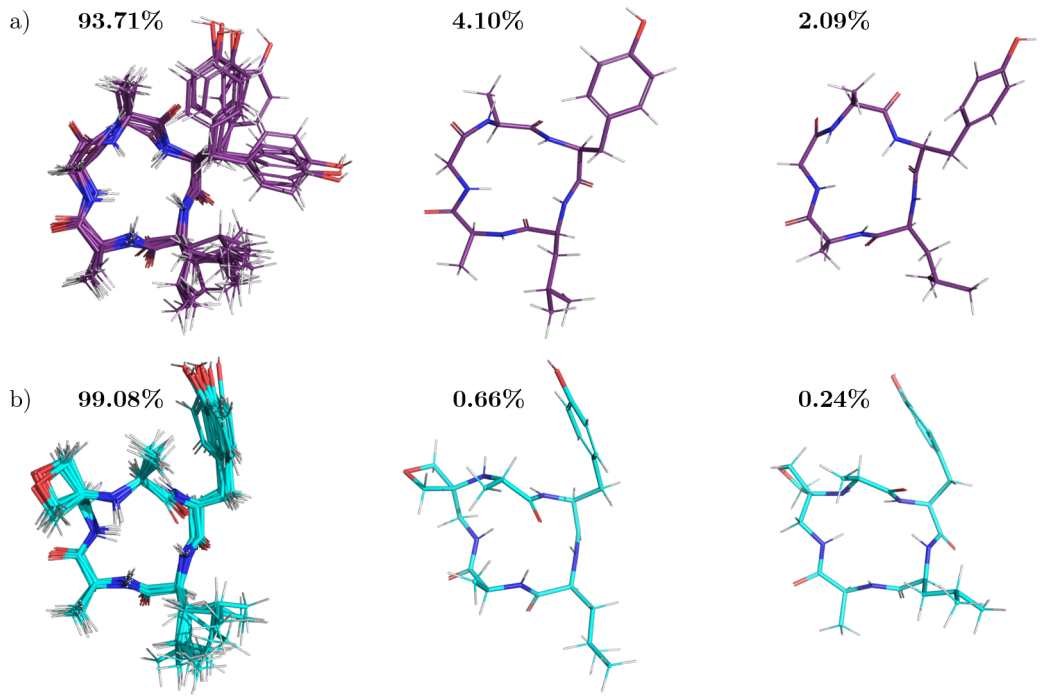


Figure 4.20: Snapshots of each of the three most populated clusters for (a) cLAGAY, and (b) cLAG$_{ox}$AY. Ten representative structures are overlaid for cluster 1, and single structures are shown for clusters 2 and 3 for each peptide. The population percentage of each cluster is indicated above.

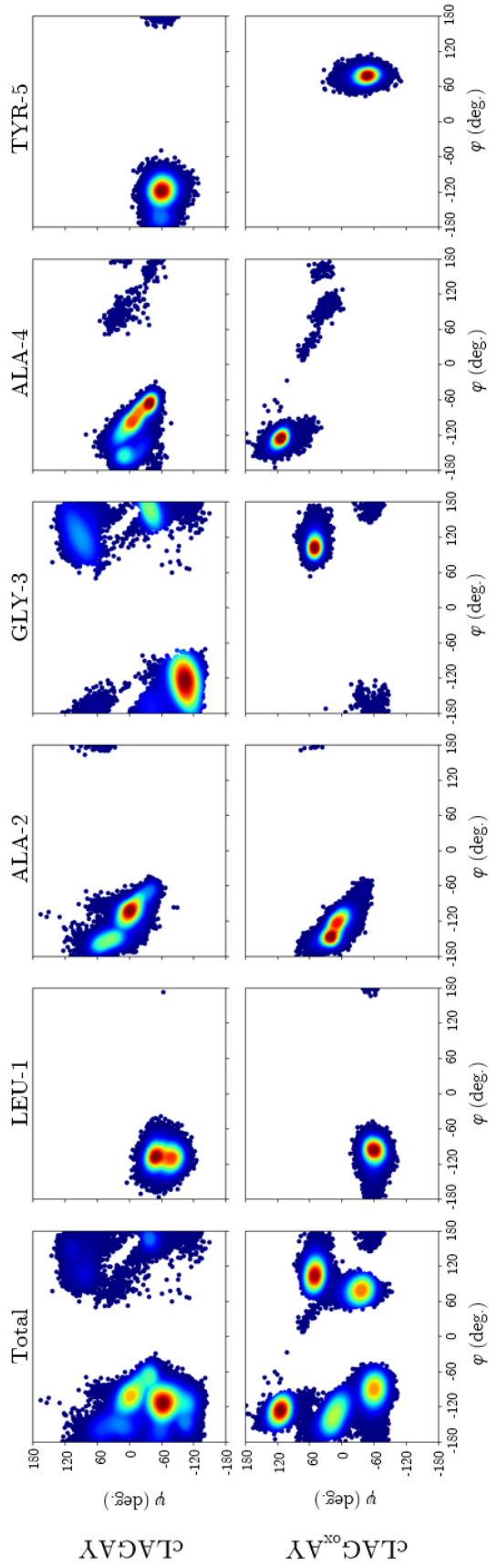


Figure 4.21: Ramachandran plots showing dihedral angles ϕ against ψ for the amino acid residues within cyclic peptides cLAGAY and cLAG$_{ox}$AY.

In order to elucidate what could be causing the increased and unexpected rigidity of cLAG$_{ox}$AY, hydrogen bonding capabilities were assessed. MD suggests that cLAG$_{ox}$AY is able to form on average ca. one intra-peptide hydrogen bond (Fig. 4.22), while cLAGAY has virtually no intra-peptide hydrogen bonding capability. However, both peptides have similar total theoretical hydrogen bonding capabilities. The hydrogen bonding pairs were assessed individually (Tables 4.6 and 4.7), and there is one dominant (86.3%) intra-peptide hydrogen bond present in the trajectory of cLAG$_{ox}$AY, corresponding to Leu1NH–Ala4O. Solvent-accessible surface area was not significantly different between the two peptides.

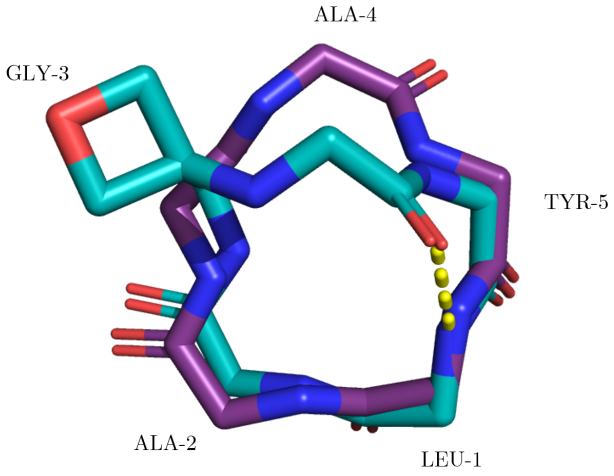


Figure 4.22: Peptide backbones of cLAGAY (purple) and cLAG$_{ox}$AY (cyan) with side-chains omitted for clarity. For cLAG$_{ox}$AY, intramolecular H-bond is indicated (yellow dots).

Table 4.6: Average number of peptide-solvent and intra-peptide hydrogen bonds formed by cLAGAY and cLAG$_{ox}$AY in DMSO.

Peptide	# of intra-peptide hydrogen bonds	# of peptide-solvent hydrogen bonds	Total # of hydrogen bonds
cLAGAY	0.002	5.034	5.036
cLAG$_{ox}$AY	0.966	4.029	4.995

Table 4.7: Occupancy of intra-peptide hydrogen bonds for cLAGAY and cLAG$_{ox}$AY in DMSO.

Peptide	Donor atom	Acceptor atom	% occupancy
cLAGAY	Gly3N	Leu1O	0.02
cLAG$_{ox}$AY	Leu1N	Ala4O	86.30
	Glx3N	Ala4O	0.03

Oxetane introduction only alters the backbone conformation of residues in proximity to the modification itself. In this case, an inversion of the amide bond between Ala4 and Tyr5 arises on oxetane introduction, which allows the structure to be stabilised by an intra-peptide hydrogen bond across the macrocycle (Fig.

4.22), corresponding to a shift of Gly3, Ala4 and Tyr5 to different regions of ϕ/ψ space. This visual representation of the backbone is in agreement with the Ramachandran plots (Fig. 4.21).

The goodness-of-fit to the restraints in the resulting structures were evaluated (Table 4.8). As with the linear peptide, the parent peptide does not experience any major restraint violations, however the strong restraint between Ala4Hα and Tyr5OH in cLAG$_{ox}$AY experiences significant deviation from its bounds. This suggests that the current force field parameters for oxetane may not be capturing the chemistry of the turn-inducing effect fully. Unrestrained simulations were also carried out, however without the NMR-derived restraints the molecules were very flexible and the simulations did not converge within a time frame that allowed for cluster analysis. Therefore, it would be prudent to further validate the oxetane parameters, for example by using MD to generate theoretical NOEs to compare to those obtained using NMR.

Table 4.8: Restraint violations (root mean squared deviation from experimental distance restraints (Å) with standard deviation) throughout simulations for cLAGAY and cLAG$_{ox}$AY.

Parameter	cLAGAY	cLAG$_{ox}$AY
All inter-residue	0.04 ± 0.05	0.25 ± 0.75
Strong	0.04	2.63
Medium	0.12 ± 0.03	0.06
Weak	0.00 ± 0.01	0.00 ± 0.01

Together, the data suggest that the oxetane modification reduces flexibility *in this specific sequence*. This may be due to a distortion in the backbone of the oxetane-modified peptide, which is stabilised by an intra-peptide hydrogen bond. As this depends on intra-peptide hydrogen bonding, the effect may be sequence-specific.

4.2.3 Structural Impact of Azetidine Modification on Linear LAGAY

The azetidine modification is of growing interest in medicinal chemistry, as it is possible to prepare functional derivatives by selective functionalisation of the azetidine nitrogen. The azetidine heterocycle is structurally similar to oxetanes, but contains an amine in place of the oxygen. We anticipated that changing the oxygen atom to nitrogen may elicit a similar turn-inducing effect, which has been beneficial in the preparation of challenging cyclic peptides. 3-aminoazetidine-3-carboxylic acid (Azt) residues containing an azetidine ring covalently bonded to

the Cα have been reported to induce a type II β-turn.[65] It is therefore of interest to assess whether azetidine-modified peptides behave in the same way as oxetane-modified peptides, and 2D NMR techniques were used to investigate whether a similar turn-inducing effect was observed.

As previously, the methyl ester of the peptides were used to eliminate any charge-charge interactions between the termini (Fig. 4.23). As with the oxetane-modified $LAG_{ox}AY$-OMe, the methyl ester of azetidine-modified LAGAY ($LAG_{az}AY$-OMe) has an increased number of medium-range NOE connectivities compared to the parent (Fig. 4.24), notably the dNN (*i, i*+2) and dαN (*i, i*+2) NOEs observed in the oxetane-modified peptide between residues Gly3 and Tyr5 (Fig. 4.13c). This suggests that the azetidine modification behaves in a similar turn-inducing manner to the oxetane in short linear peptides.

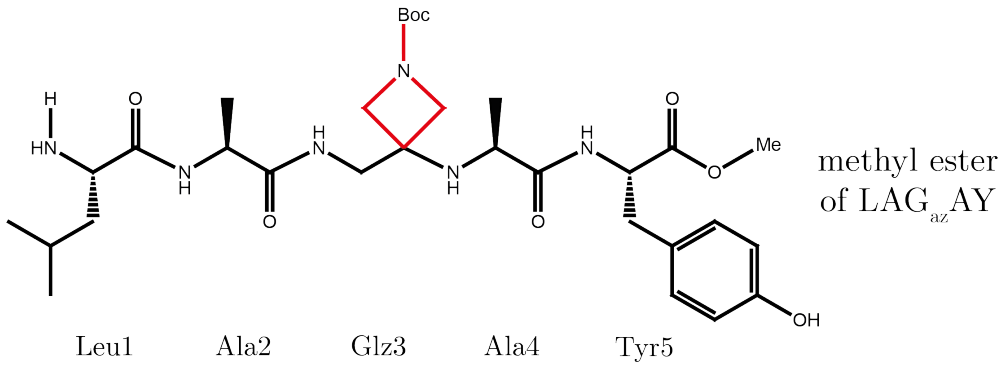


Figure 4.23: Structures of methyl ester of azetidine-modified peptide $LAG_{az}AY$-OMe. The azetidine is highlighted in red, and contains the Boc protecting group.

4.3 Conclusions

A major challenge with peptide-based drugs is their poor resistance to proteases. Peptide modifications have been used to reduce vulnerability to exo-peptidases, for example with incorporation of N-terminal D-amino acids or C-terminal reduction of the carboxylic acid into the corresponding alcohol.[45] Head-to-tail cyclisation is particularly advantageous, as it confers resistance to both endo- and exo-peptidases.

Cyclic peptides remain promising as potential therapeutics, yet they are challenging to synthesise. Peptide modification provides an avenue for the increased efficiency of macrocyclisation, and oxetane modification has been shown to bring the termini of a linear tripeptide closer together, potentially increasing the efficiency of cyclisation.[7] In order to understand the effect of oxetane modification on both linear and cyclic peptides, we used a combination of biophysical techniques and molecular dynamics simulations.

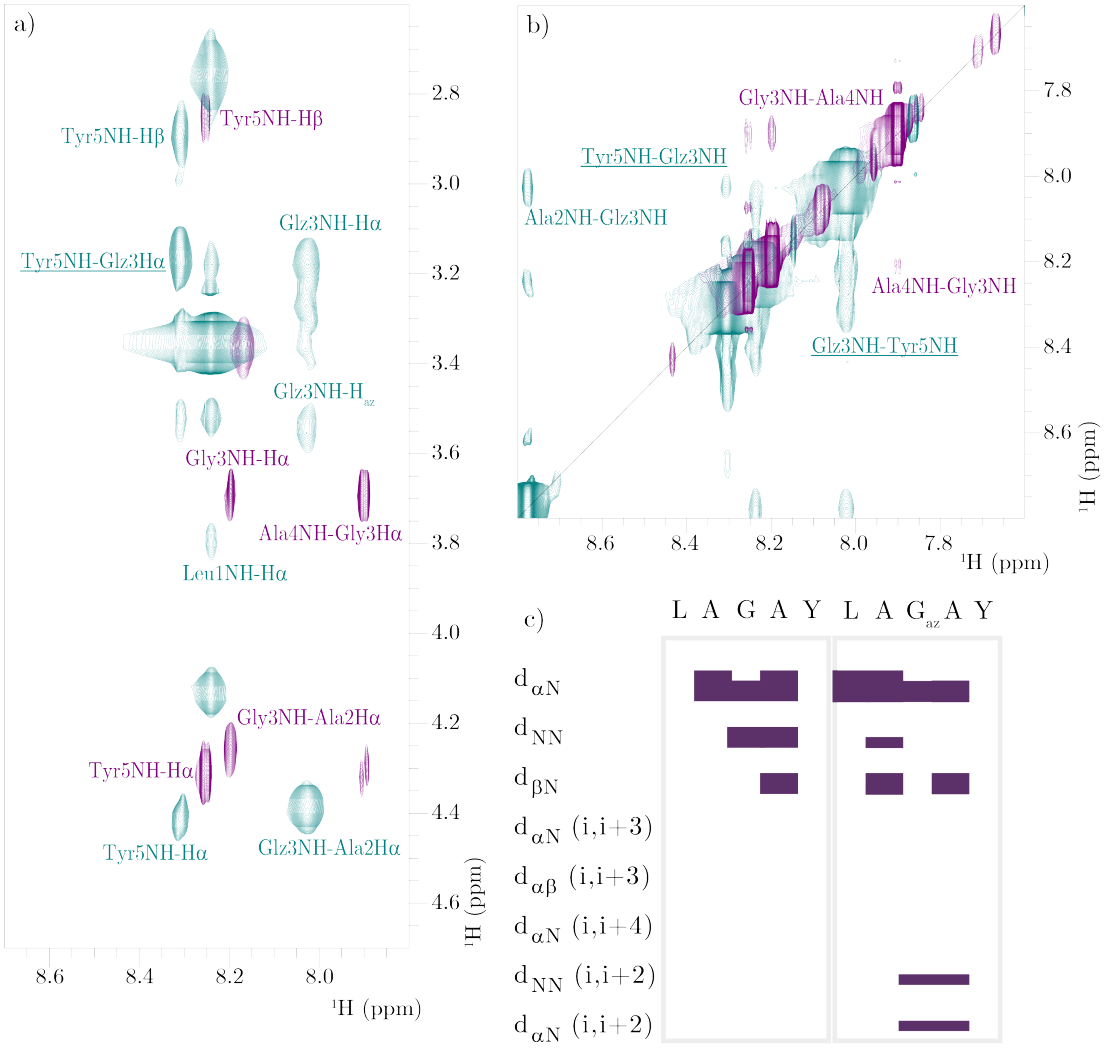


Figure 4.24: (a) Fingerprint region and (b) amide region of NOESY spectra of $LAG_{az}AY$-OMe (cyan) and LAGAY-OMe (purple) collected with a mixing time of 400 ms for samples containing an unknown concentration of peptide solubilised in d_6-DMSO. Both spectra were collected at 298 K at a field strength of 700 MHz. Medium-range NOEs are underlined. (c) NOE-connectivity maps of $LAG_{az}AY$-OMe and LAGAY-OMe based on NOE spectra collected with 250–800 ms mixing times. Each sequential and medium-range coupling is indicated on the left, and a bar is used to join the observed coupled residues.

Solution-state NMR shows the presence of medium-range NOEs in oxetane-modified linear LAGAY, which are absent in the unmodified parent peptide. These are observed in both the zwitterionic and methyl ester forms of the peptide, and are indicative of the formation of a turn.[92,132] Additionally, MD simulations with NMR-derived restraints performed on LAGAY-OMe and $LAG_{ox}AY$-OMe suggest that $LAG_{ox}AY$-OMe was much more flexible than the unmodified parent. In fact, $LAG_{ox}AY$-OMe spends over 10% of its time in a conformation where the termini are within 5 Å, while the parent peptide spends a tiny fraction of its time in a similar conformation.

This agrees with previous observations, which suggested that oxetane-modified linear tripeptides were more flexible than their unmodified counterparts,[7] and have the termini closer together more frequently. This may explain why oxetane-modified peptides have an increased efficiency of macrocyclisation, as confirmed experimentally by comparing the yields of cyclic peptides. The yields of cyclic peptides are consistently higher following oxetane modification, by around 20% for the sequence LAGAY. The increase in yield following oxetane modification is substantially higher than that following other common modifications such as N-methylation. The largest improvement in the macrocyclisation efficiency is observed when the oxetane modification is located in the centre of the precursor backbone.[92]

A similar effect has been observed in azetidine-modified peptides, and it was shown using 2D NMR techniques that addition of an azetidine results in a similar pattern of medium-range NOEs, suggesting azetidine modification also has a turn-inducing effect in short linear peptides. Importantly, unlike the oxetane modification, the azetidine modification is robust under acidic conditions which are required to deprotect the sidechains of the amino acids in the macrocycle.[41] Additionally, azetidines provide a route for functionalisation, as the azetidine ring itself can be modified to provide useful derivatives, for example by the attachment of dyes or radionucleotides (for imaging),[75,169] drug conjugates (for receptor targeting)[76] or additional peptide sequences (to aid cell penetration).[77]

As discussed, a cyclic backbone is a highly desirable trait when selecting a drug molecule. Thus, it is important to characterise the effect of peptide modification on cyclic peptides. The effect of oxetane modification of a cyclic pentapeptide, cLAGAY, was assessed. CD spectra show clear differences between modified and unmodified peptides, and our data appears similar to that for other cyclic peptides characterised using CD.[164,167] Both cLAGAY and $cLAG_{ox}AY$ were fully assigned using 2D NMR techniques, further illustrating distinct differences between the two peptides. Using MD simulations with NMR-derived restraints, the data suggest that the oxetane rigidifies this peptide, and the structure is stabilised by the for-

mation of an additional intrapeptide hydrogen bond.[92] It is difficult to directly compare the cluster analysis results for the linear and cyclic LAGAY obtained from the MD simulations due to a different cut-off used. This was selected because of the rigidity of the cyclic peptides compared to their linear precursors (the reduction in flexibility caused by macrocyclisation is well-documented).[170–172] However, it is possible to compare the structures between the modified and unmodified cyclic peptides. There is a change in the conformational space sampled for all five residues for the modified peptide, which corresponds with the rigidity observed in the cluster analysis.

Overall it appears that oxetane-modification is a useful modification in the production of cyclic peptides, and it is important that future directions of this work confirm that oxetane-aided cyclisation does not reduce any biological activity present in the linear precursor. Cyclisation can lead to a loss of activity if the reduced flexibility fixes the compound in an inactive conformation,[45] but in many cases cyclisation has been shown to increase biological activity.[46–48]

Chapter 5

Oxetane Modification on Helical Peptides

5.1 Introduction

Protein function is directly affected by structure, and in biology secondary structural motifs including α-helices and β-sheets are widely recognised as important. The α-helix was first characterised in 1951 and is one of the most common secondary structure conformations of natural proteins.[173] A typical α-helix consists of amino acids arranged in a helical structure with a characteristic repeating hydrogen bond pattern, where each carbonyl oxygen forms a hydrogen bond with a backbone amide hydrogen four residues away (Fig. 5.1).

In an α-helix, each residue is related to its neighbour by a translation of 1.5 Å along the helix axis and a 100° rotation, giving 3.6 residues per turn of the helix. The screw-sense of an α-helix can be right- or left-handed, but most α-helices found in nature are right-handed, as α-helices composed of L-amino acids in a left-handed conformation are destabilised due to steric crowding of the sidechains and the carbonyl groups.[9] In a right-handed α-helix, all the amino acids have negative ϕ and ψ angles, typically around -60° and -40° respectively. The α-helix plays a multitude of biological roles, including structural scaffold,[174] cell signalling initiator,[175] and membrane curvature sensor[176] amongst others. Additionally, the ability of α-helices to mediate protein-protein interactions (PPIs) makes them desirable therapeutic candidates.[177–179] Thus, it is important to consider how peptide modifications can affect secondary structure.

Backbone modifications can alter the biological properties of α-helices. For example, substitution of α- to β-amino acids has been shown to improve the pharmacokinetic properties of parathyroid hormone derivatives.[180] Other modification methods have been used to stabilise or induce helicity, such as the incorporation

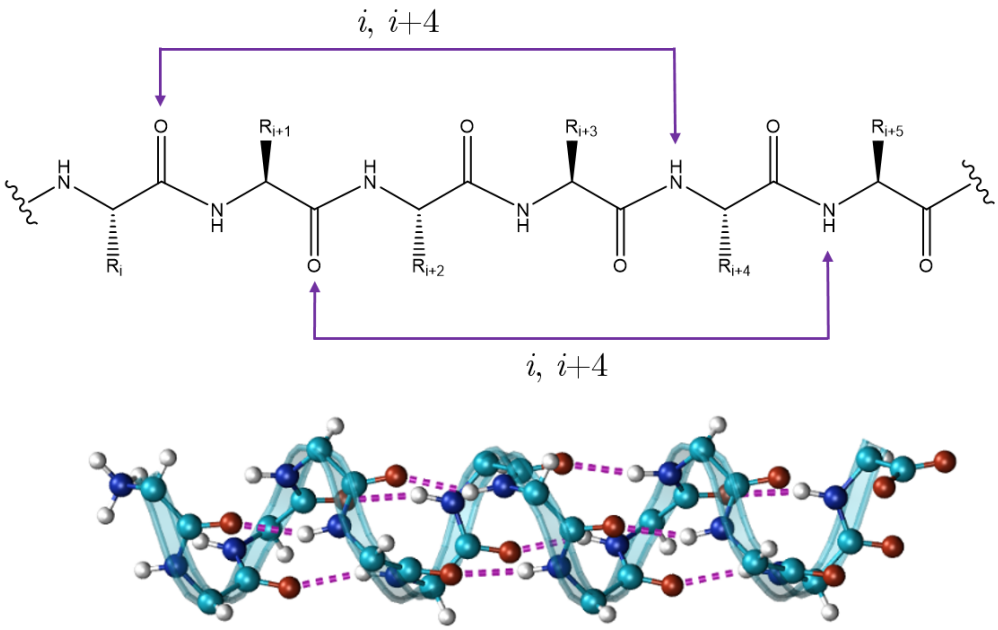


Figure 5.1: Hydrogen bonding scheme (top) and a ball-and-stick depiction (bottom) of an α-helix. In both representations, *i, i*+4 hydrogen bonds are shown by purple lines.

of Cα-alkylated amino including 2-aminoisobutyric acid,[30,181] or by stapling short to mid-size peptides by constraining them with a synthetic brace composed of non-natural amino acids or hydrocarbons.[30]

Short alanine-based peptides form highly stable α-helices in water,[19,182,183] and these model systems have been used to assess how amino acid substitution affects helix stability.[19,182] Alanine-based peptides have also previously been used to investigate the impact of peptide backbone modifications on α-helix structure and stability. For example, thioamide substitution at either a central or N-terminal position of an alanine-based model peptide has a destabilising effect on the helix, with the impact greatest when the modification is in the centre of the peptide.[184]

Oxetanes have previously attracted significant attention in medicinal chemistry,[53] as substituting *gem*-dimethyl or carbonyl groups with an oxetane motif can trigger profound changes in the stability, lipophilicity, solubility and conformational preference of small molecules. Despite the fact that the C-O distance in an oxetane ring is larger than in a carbonyl group (2.1 vs. 1.2 Å),[53] the two motifs have a similar lone pair arrangement or hydrogen bonding capabilities,[57,58] thus oxetane modification could result in minimal disruption to helicity whilst modulating the pharmacokinetic properties of the molecule.

This work builds on previous results in which a model helical peptide, Ac-[KAAAA]$_3$-KGY-NH$_2$, was selected to test the structural impact of oxetane mod-

ification at two sites in an α-helix. Specifically, oxetane modification was introduced at the N-terminus and in the centre of a helical region (Table 5.1). These two sites were selected based on previous research involving modifications of α-helical structures, which suggest that modifications are better tolerated in the terminal regions of α-helices due to their lower intrinsic helical content.[182,184] The synthesis of these peptides was outside of the scope of this thesis, and is described elsewhere.[10] A previous report generated by our collaborators used circular dichroism (CD) data to show that addition of an oxetane heterocycle impacts helical content dramatically (see Fig. 5.2).[9,93] The greatest disruptive effect was observed when methanol (which promotes the formation of helices by strengthening hydrogen bonds)[165] was used as a solvent (Fig. 5.2b). The percentage helicity of the oxetane-containing peptides used are displayed in Table 5.2, below. Oxetane modification was performed at position 3 (henceforth referred to as N-term. Alx, for N-terminal oxetane-modified alanine, and N-term. Glx, for N-terminal oxetane-modified glycine) and position 8 (central Alx).

Table 5.1: Sequences of the peptides used. A_{ox} indicates the presence of an oxetane-modified alanine and G_{ox} indicates the presence of an oxetane-modified glycine.

Sequence	Peptide	Modified position
Ac-KAAAAKAAAAKAAAA-KGY-NH_2	Parent	N/A
Ac-KAAAAKAA_{ox}AAKAAAA-KGY-NH_2	Central Alx	8
Ac-KAA_{ox}AAKAAAAKAAAA-KGY-NH_2	N-term. Alx	3
Ac-KAG_{ox}AAKAAAAKAAAA-KGY-NH_2	N-term. Glx	3

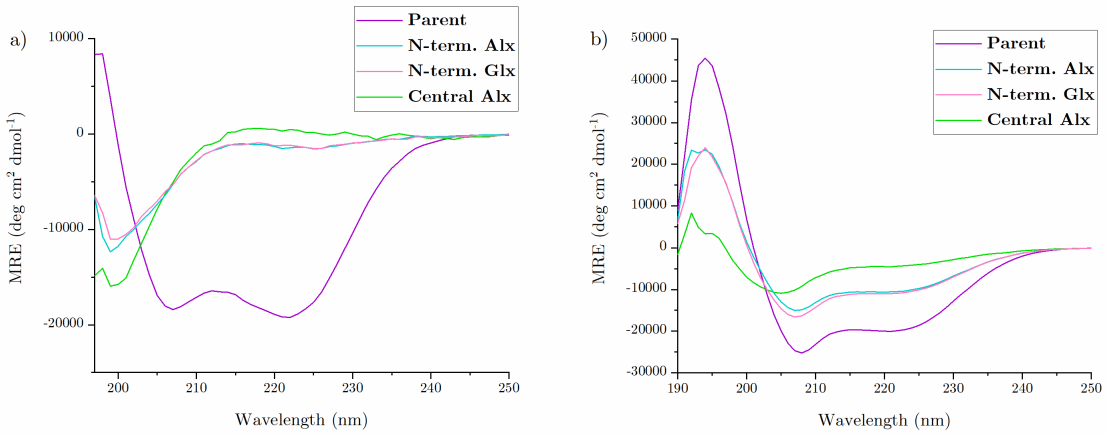


Figure 5.2: Circular dichroism spectra acquired for 0.1 mg/mL solutions of each peptide shown in Table 5.1 in (a) 10 mM potassium phosphate buffer (pH 7.0) with 1M NaCl, and (b) 100% methanol. Negative peaks at 208 and 222 nm indicate the presence of α-helical structure.

Table 5.2: Secondary structure of the peptides in MeOH at 0 °C estimated by circular dichroism. Helical content was calculated using DichroWeb (Selcon3, Ref. set 4).[125] Reproduced from Beadle.[9]

Peptide	Helicity (%)
Parent	57
Central Alx	15
N-term. Alx	34
N-term. Glx	37

In this chapter, work is summarised that reveals the long-range and short-range molecular details of how the oxetane modification destabilises α-helices. Two-dimensional NMR methods were used to explore the impact of incorporation of an oxetane-modified glycine at a single site in an α-helical structure, and observed that oxetane modification was poorly tolerated. Experimental NMR data were complemented with computational methods, and we used work required to unwind the peptides as a measure which correlates with helical content observed using CD. The simulations provide details of the structural impact surrounding the incorportation of an oxetane-modified alanine at two different sites, and an oxetane-modified glycine at a single site, and could be used to guide future design of chemically modified peptides.

5.2 Results and Discussion

5.2.1 NMR Spectroscopy Shows a Reduction in Helicity Upon Oxetane Modification

The parent peptide and the N-term. Glx peptide were analysed using solution state NMR spectroscopy. ^{1}H–^{1}H TOCSY and NOESY data were acquired for both peptides in 80% MeOD-d_4, and assignment was attempted. As expected, the repetitive nature of the sequence prevented full sequential assignment. However, all four Lys residues in the peptide were well-resolved and the unique Gly and Tyr residues were readily assigned, as was the oxetane-modified Glx in the N-term. Glx peptide. NOEs were used where possible to sequentially assign additional residues, but signals from the majority of the Ala residues were heavily overlapped and identification of unique *i, i*+4 NOE correlations (characteristic of α-helix formation) was not possible. NMR spectra for both peptides are shown in Figs. 5.3 and 5.4 (fingerprint regions) and 5.5 (NH-NH region) and resulting assignments for both peptides are shown in Table 5.3. As Hα chemical shifts are highly sensitive to changes in secondary structure,[132,185] these were used as chemical probes for

changes in chemical environment at regular points along the length of the peptide chain.

The Hα chemical shifts of residues 15–18, distant from the oxetane modification, remained unchanged to within 0.01 ppm upon introduction of the oxetane.

In contrast, upon introduction of the oxetane modification at position 3, a significant downfield shift of $\geq$0.21 ppm for the Lys 1 Hα peak was observed (from 4.02–4.05 ppm in the parent peptide to 4.26 ppm in N-term. Glx). For one additional Lys residue (either Lys 6 or Lys 11), a similar downfield shift of $\geq$0.17 ppm was also observed from 4.02—4.05 ppm in the parent to 4.22 ppm in N-term. Glx. A downfield shift for an Hα proton signal of 0.1 ppm or more suggests a change in secondary structure from an α-helix to an unstructured (random coil) chain, reflecting the differences in ^{1}H chemical environment and hydrogen bonding patterns in regions of helical structure.

Table 5.3: Partial ^{1}H chemical shift assignments for parent and N-term. Glx peptides solubilised in 80% MeOD-d_4 + 20% H_2O to a final peptide concentration of 2 mM. Due to the repetitive nature of the sequence, sequential assignment was not possible for the majority of residues. Unique residues were readily identified and sequentially assigned, and NOE data was used where possible to sequentially assign additional residues. Lys residues followed by multiple sequence positions have been assigned by residue type, but not by position, so all possible sequence positions are noted.

Parent	Chemical shift (ppm)			
Residue	NH	Hα	Hβ	others
Lys (Lys 1, 6, 11)	8.70	4.05	1.86	Hγ 1.50; Hδ 1.73; Hε 2.97
Lys' (Lys 1, 6, 11)	8.24	4.02	2.00	Hδ 1.70; Hε 2.92
Lys" (Lys 1, 6, 11)	8.19	4.03	1.99	Hδ 1.70; Hε 2.92
Ala15	8.10	4.18	1.55	
Lys16	7.66	4.26	1.96	Hγ 1.53; Hδ 1.68; Hε 2.95
Gly17	7.99	4.00		
Tyr18	7.96	4.52	2.85/3.11	
N-term. Glx	Chemical shift (ppm)			
Residue	NH	Hα	Hβ	others
Lys 1	8.40	4.26	1.84	Hγ 1.49; Hδ 1.71; Hε 2.97
Ala2	8.46	4.30	1.42	
Glx3	8.22	3.65	4.12	
Ala4	8.72	4.25	1.49	
Lys (Lys 6, 11)	8.78	4.14	1.88	Hγ 1.57; Hδ 1.73; Hε 2.96
Lys' (Lys 6, 11)	8.14	4.22	1.97	Hδ 1.70; Hε 2.94
Ala15	8.09	4.19	1.55	
Lys16	7.68	4.26	1.95	Hγ 1.50; Hδ 1.69; Hε 2.94
Gly17	8.01	4.00		
Tyr18	7.97	4.53	2.85/3.12	

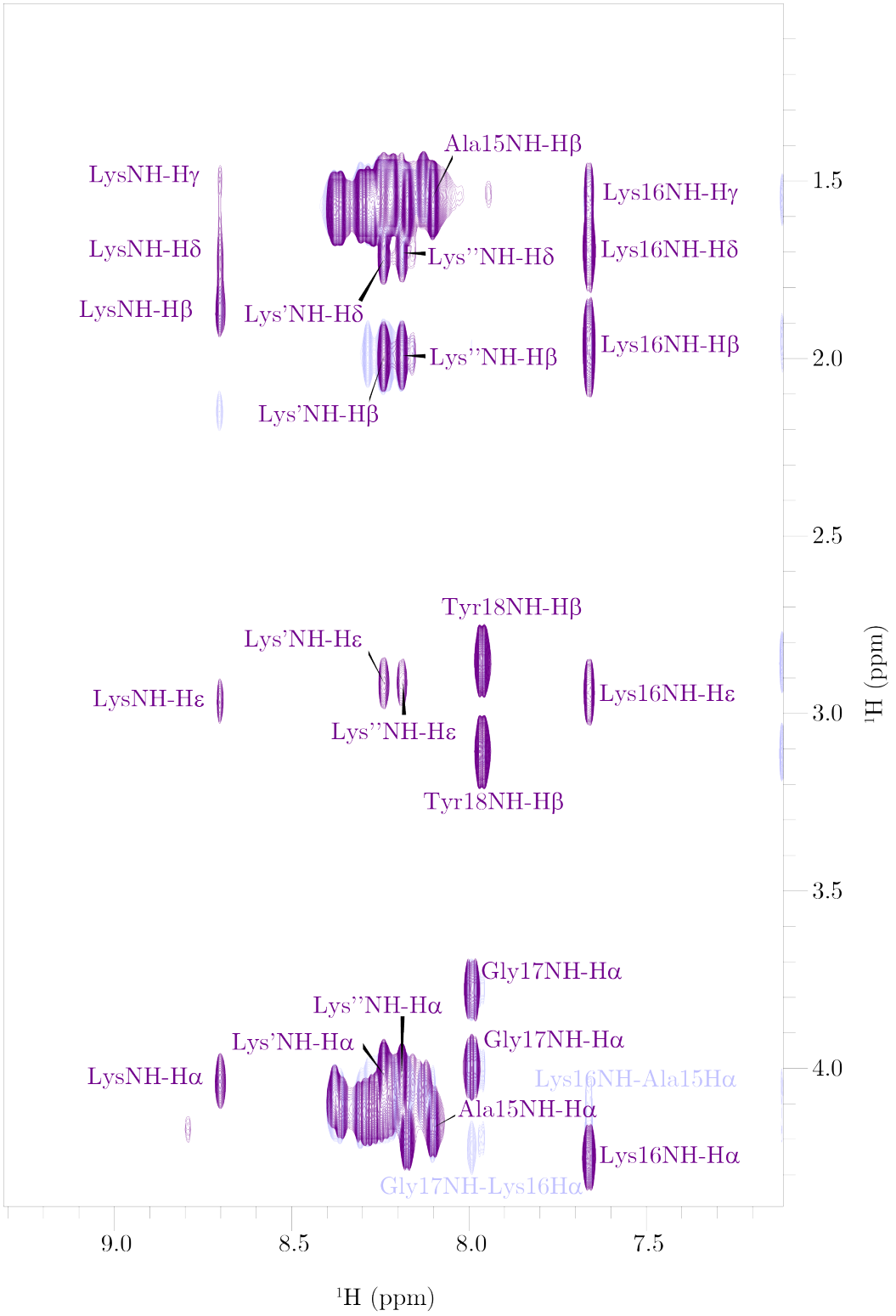


Figure 5.3: Fingerprint region of the TOCSY spectrum (purple contours, 140 ms mixing time) for the parent peptide solubilised to a final concentration of 2 mM in 80% MeOD-d_4 + 20% H_2O. The TOCSY spectrum is overlaid onto the NOESY spectrum (lavender contours, 150 ms mixing time). Non-sequential Lys assignments are denoted Lys, Lys' and Lys". Both spectra were collected at 283 K at a field strength of 700 MHz.

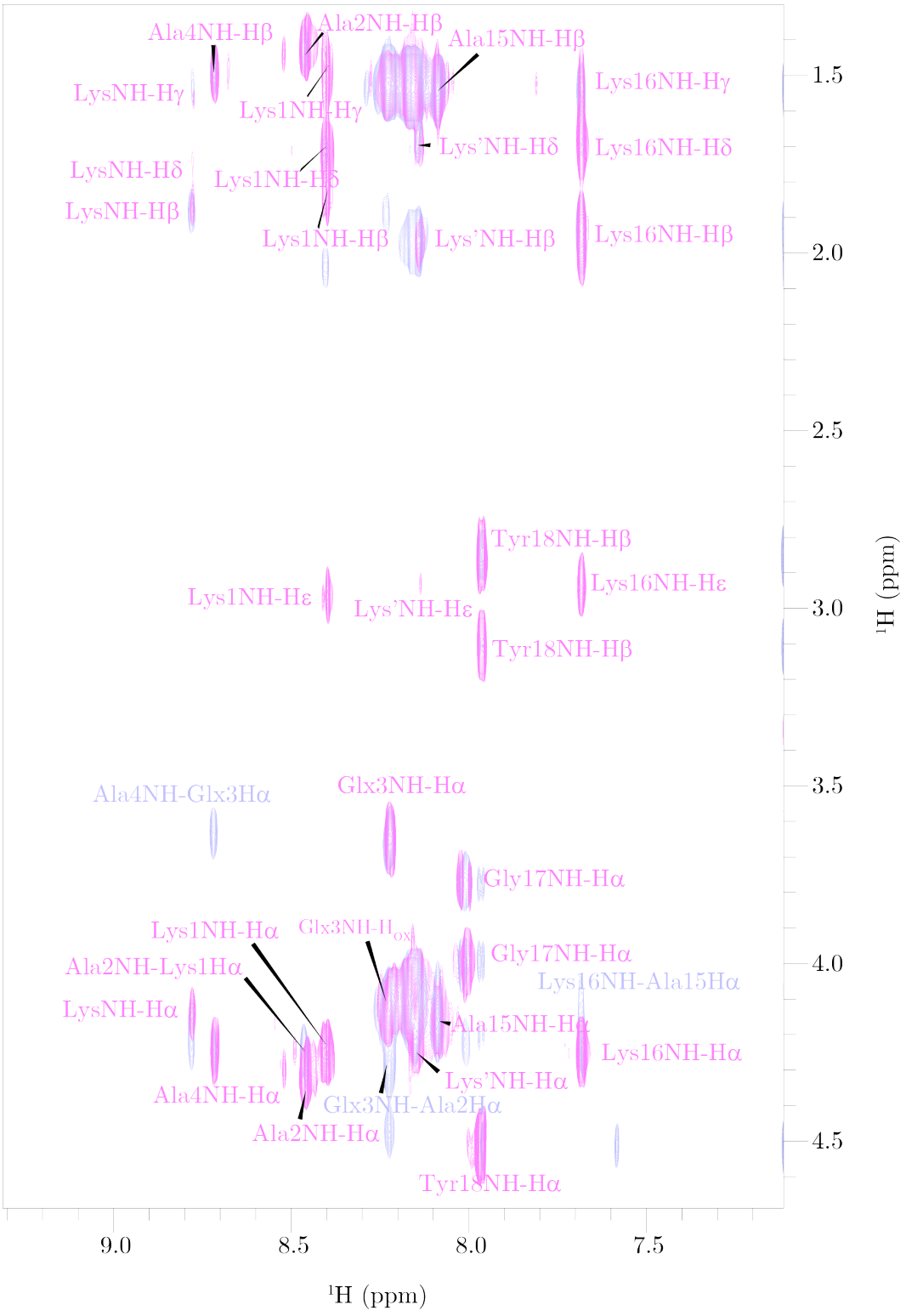


Figure 5.4: Fingerprint region of the TOCSY spectrum (pink contours, 140 ms mixing time) for the N-term. Glx peptide solubilised to a final concentration of 2 mM in 80% MeOD-d_4 + 20% H_2O. The TOCSY spectrum is overlaid onto the NOESY spectrum (lavender contours, 150 ms mixing time). Non-sequential Lys assignments are denoted Lys and Lys'. Both spectra were collected at 283 K at a field strength of 700 MHz.

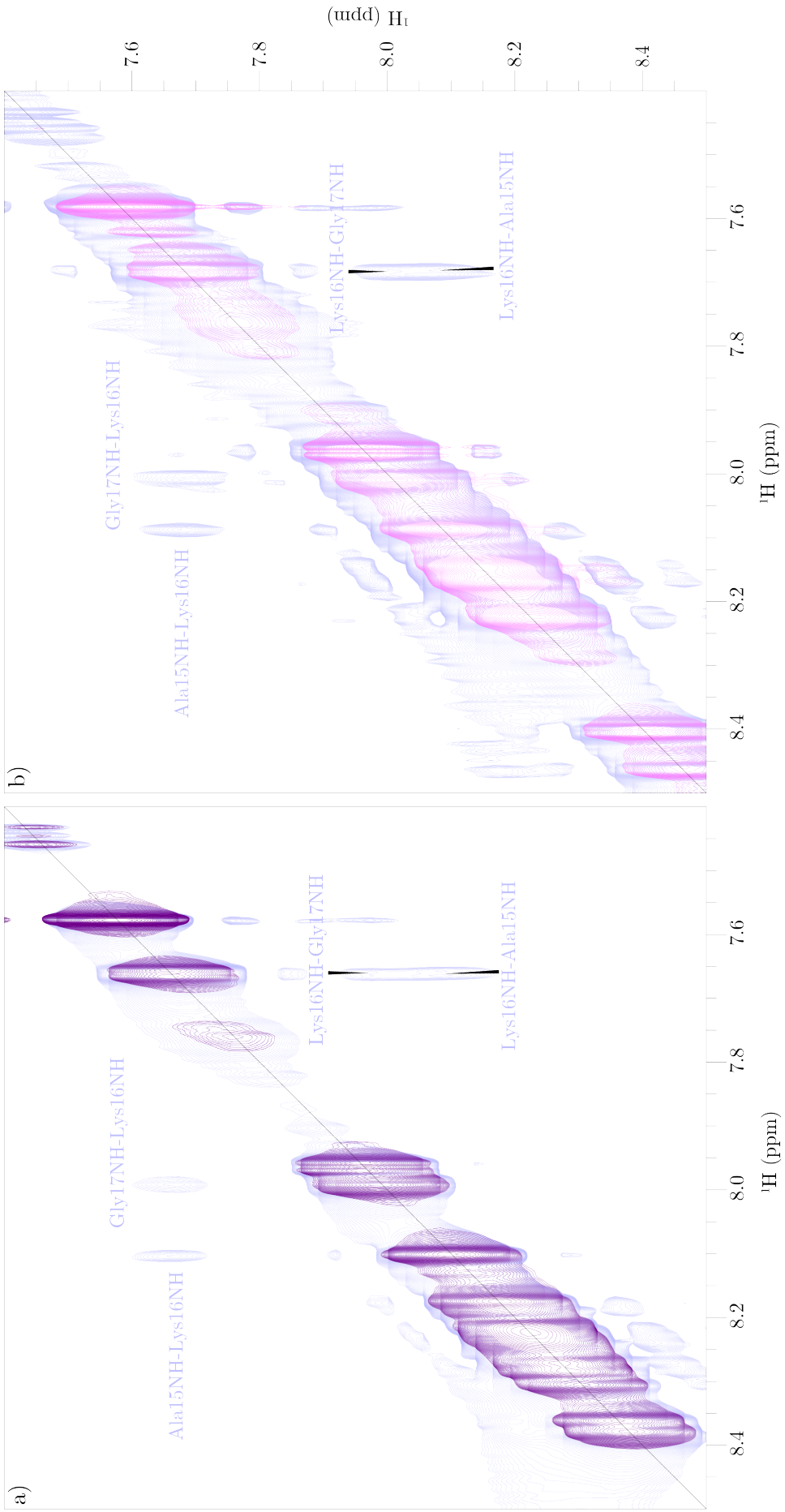


Figure 5.5: Amide region of TOCSY and NOESY (lavender) spectra for (a) parent peptide (purple) and (b) N-term. Glx peptide (pink). For both peptides, the NH-NH NOE correlations support the sequential assignment of residues 15–17.

This relationship, alongside the CD data described previously,[9,93] suggests that Lys 1, 6 and/or 11 are all in an α-helical environment in the parent peptide, whilst only one Lys (either Lys 6 or 11) is in an α-helical environment in the N-term. Glx peptide. To gain further insights into the impact of the oxetane modification on the structure and stability of the helical peptides, we decided to use a predominantly computational approach.

5.2.2 Brute Force Simulations Are Not Suitable for This Model System

5.2.2.1 Selection of Force Fields

There is some evidence to suggest that certain force fields, including CHARMM, stabilise helices too much.[186] Therefore, we decided to compare how our model system behaved using three different force fields: GROMOS96 54a7, AMBER99SB-ILDN and CHARMM27. In order to keep the simulations as simple as possible, the parent (unmodified) helical peptide was simulated using each force field for 500 ns using zwitterionic ends and protonated lysines in a box of 6942–6960 water molecules. Charges were countered using chloride ions.

The peptide was built in Avogadro 1.2.0[139] with ϕ and ψ angles corresponding to an idealised α-helix. It was then briefly subjected to a steepest descent energy minimisation in Avogadro using the Universal Force Field.[144] The resulting structure was used for simulations using all three force fields using GROMACS 5.1.4.[137] The initial structures were solvated in water and subjected to 50000 steps of energy minimisation using the steepest descent algorithm. For each force field, the peptide was then relaxed by performing 50000 steps of MD simulation in the NVT ensemble and 50000 steps of simulation at 300 K and 1 bar in the NPT ensemble to equilibrate the temperature and density of the system respectively. The final production run was carried out, using each force field, for 500 ns in the NPT ensemble.

Ramachandran plots (Fig. 5.6), showing the distribution ϕ and ψ backbone dihedral angles of the entire peptide (averaged over the last 50 ns of simulation trajectory), indicate large differences between the three force fields. The plot for the simulation using CHARMM27 (Fig. 5.6a) indicate a predominantly right-handed α-helical conformation (negative ϕ, negative ψ), with only minor sampling of the rest of the conformational space. The simulation using AMBER99SB-ILDN (Fig. 5.6b) results in sampling of the ϕ/ψ corresponding to β-sheets (negative ϕ, positive ψ), as well as some left-handed α-helical content (positive ϕ, positive ψ), and some minor right-handed α-helical conformations. The simulation

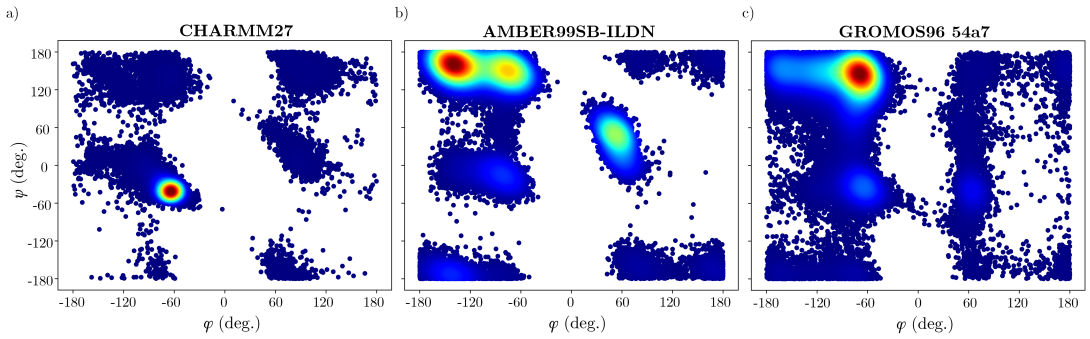


Figure 5.6: Ramachandran plots for the parent helical peptide, simulated for 500 ns using (a) CHARMM27, (b) AMBER99SB-ILDN, and (c) GROMOS96 54a7. There are clear differences between the three plots. Data shown represents the period of 450–500 ns of the trajectory.

using GROMOS96 54a7 (Fig. 5.6c) results in a plot that appears to be a middle ground between CHARMM27 and AMBER99SB-ILDN—large amounts of β-sheet content, with some right-handed α-helix present.

As we expect this sequence to be predominantly helical, based on the experimental data, we selected CHARMM27 moving forward. Although CHARMM27 may be overly stabilising this helical structure, it would appear that AMBER99SB-ILDN and GROMOS96 54a7 may not be stabilising the helix enough for this model system. An additional benefit of selecting CHARMM27 is that this force field already contains compatible parameters for the oxetane-modified amino acids (previously described by Powell et al.[7]).

5.2.2.2 CHARMM Simulations

Following equilibration as previously described, the helical peptides were simulated for a total of 2 μs in the *NPT* ensemble, which was not sufficient for helical unwinding to occur. In fact, the peptides switched back and force between two states—fully wound (Fig. 5.7a) and N-term. slightly unwound (Fig. 5.7b). The root mean squared deviation compared to the starting structure of the peptide does not stabilise during the simulation, and instead reflects the switching between the two states. Longer tests of 10 μs with uncharged lysine sidechains suggested that complete helical unwinding would require milliseconds to seconds of simulation time, as like the shorter tests the peptide switched between fully wound and with the N-terminus slightly unwound. Thus, brute force MD simulations were deemed to be not computationally feasible.

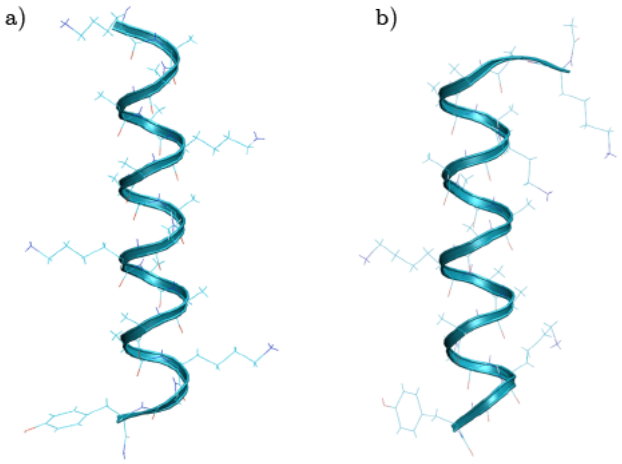


Figure 5.7: Partial unwinding of the unmodified helical peptide during long brute force MD simulations. The peptide alternates between two states, (a) fully helical, and (b) N-terminus slightly unwound. This process was deemed to be computationally unfeasible.

5.2.3 Steered MD Simulations Can Be Used To Unwind Helical Peptides

Steered molecular dynamics (SMD) is analogous to single molecule spectroscopy, in which a molecule is fixed at one end and pulled at another, and has previously been used to aid unwinding of helical structures,[121,122] as it allows unwinding events to occur on computationally feasible time scales. SMD simulations were carried out on the four helical peptides in methanol. Each peptide was built using Avogadro 1.2.0[139] with ϕ and ψ angles corresponding to an idealised α-helix. The carbonyl at position 3 or 8 for the N-term. Alx/Glx and central Alx peptides respectively was substituted with the oxetane group. The parent, unmodified peptide was also subject to simulation. All four peptides had N-terminal acetyl and C-terminal amide caps added (to better represent the structures used for CD and NMR analysis), before being briefly subjected to a steepest descent energy minimisation in Avogadro using the Universal Force Field.[144] All simulations were performed using the GROMACS 5.1.4 simulation package.[137] using the CHARMM27 force field[97] with modifications for the oxetane ring[7] for the peptides and the CHARMM General Force Field (CGenFF) for methanol.[187] Each peptide was simulated ten times for 40 ns at 300 K (each repeat had different initial co-ordinates). The peptides were anchored at the N-terminus with a dummy spring pulling the C-terminus. However, the parameters used in steered MD are not well-characterised, and first required extensive testing and optimisation.

5.2.3.1 Parameter Optimisation

Simulation parameters for the steered MD were selected following a series of short (1–10 ns) tests in which spring constants of 100 to 10 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and pull speeds of 0.02 to 0.00005 $\mathrm{nm\,ps^{-1}}$ were assessed. These initial values were selected based on other work in the literature.[120] Following observations that pull speeds of 0.02 to 0.001 $\mathrm{nm\,ps^{-1}}$ resulted in rapid helix unwinding even in short tests, further repeats showed that with slow pull speeds (0.0005 $\mathrm{nm\,ps^{-1}}$), use of a higher spring constant results in less error between repeats as estimated from the graphs but more overall noise in the force extension curves (Fig. 5.8a and b). Thus a spring constant of 25 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and a pull speed of 0.00025 $\mathrm{nm\,ps^{-1}}$ were selected, as these parameters allowed for slow, gentle unwinding over the course of a longer simulation, while producing force extension curves with minimal noise. Reducing noise was deemed to be more important than reducing error between repeats, as at higher spring constants the signal-to-noise ratio was low and it was challenging to determine which peaks corresponded to unwinding events.

Slower pulling rates of 0.0001 and 0.00005 $\mathrm{nm\,ps^{-1}}$ resulted in the production of similar trajectories and force extension curves (Fig. 5.8c), thus the faster pull speed was selected to reduce computational expense while maintaining integrity of the data. We also compared pulling from the N-terminus and the C-terminus in a series of short tests, and found that with these parameters, there were no significant differences between the two, therefore we opted to use the Cα atom of Lys1 as an immobile reference, and pulled the Cα atom of Tyr18 along the z-axis. Under these conditions, each peptide fully unwinds following 18–23 ns of simulation time.

5.2.3.2 Work Required to Unwind Oxetane-Modified Helices Reflects a Weakening of the Fold

The stability of a protein fold is reflected in the energy that is required to unfold the protein. We calculated the work to pull the Cα atom of Tyr18 of each helix by 5 nm by integrating the force extension curves produced during the SMD simulations. Fig. 5.9 shows the amount of work required to unwind each helix averaged over the 10 repeats. Introduction of oxetane-modified alanine at position 3 (N-term. Alx) significantly decreases the amount of work required to unwind the helix 5 nm compared to the parent (from 127.21 $\mathrm{kJ\,mol^{-1}}$ to 111.30 $\mathrm{kJ\,mol^{-1}}$). Similarly, introduction of Glx at position 3 (N-term. Glx) decreased the amount of work required compared to the unmodified parent peptide by 10.4 $\mathrm{kJ\,mol^{-1}}$ (total work = 116.8 $\mathrm{kJ\,mol^{-1}}$). The difference between the work required to unwind the N-term. Alx and N-term. Glx peptides 5 nm is statistically significant ($p < 0.02$, indepen-

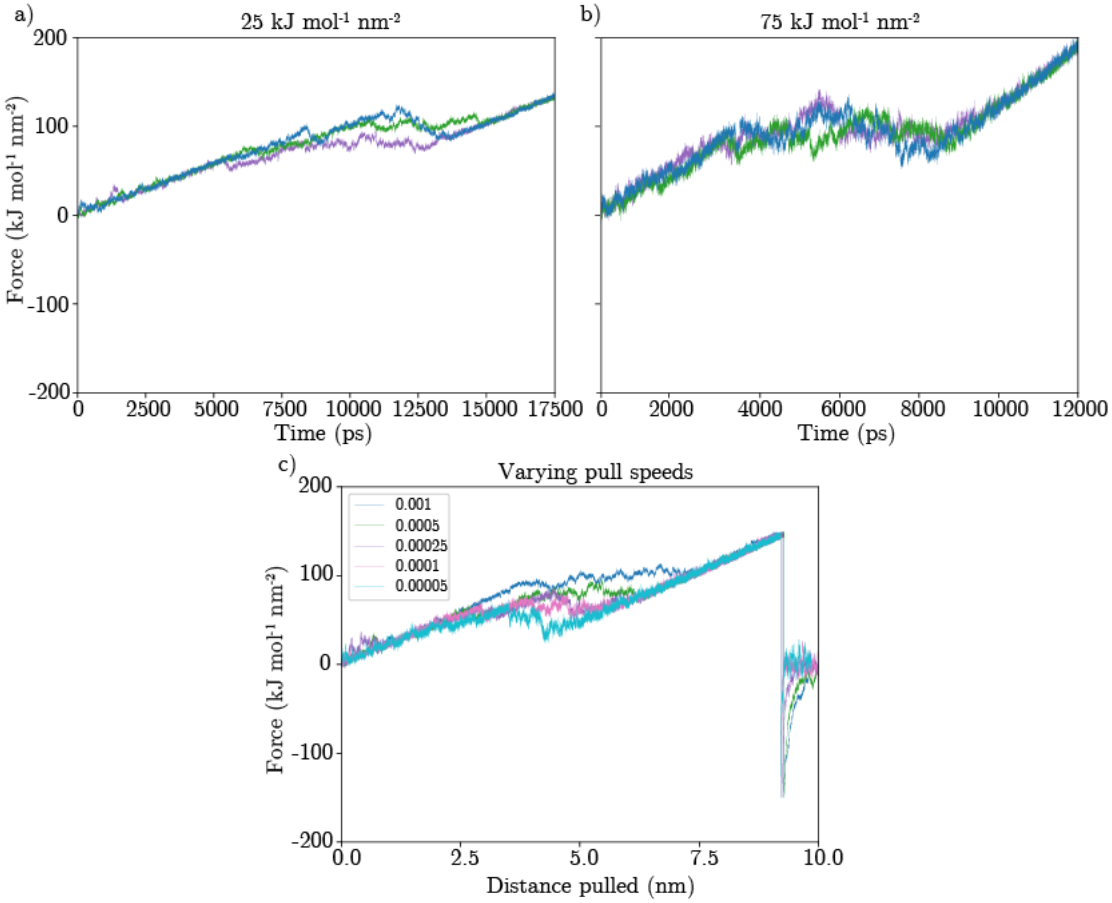


Figure 5.8: Representative force extension curves showing three repeats with different starting co-ordinates for spring constants (a) 25 kJ mol^{-1} nm^{-2} and (b) 75 kJ mol^{-1} nm^{-2}, while maintaining the pull speed at 0.0005 nm ps^{-1}, and (c) different pull speeds (0.001 to 0.00005 nm ps^{-1}) with a constant spring constant of 25 kJ mol^{-1} nm^{-2}. Based on these tests, a spring constant of 25 kJ mol^{-1} nm^{-2} and a pull speed of 0.00025 nm ps^{-1} were selected.

dent samples t-test), suggesting that the structure of the N-term. Glx peptide is moderately more stable. Oxetane incorporation at the central residue, position 8 (central Alx), had the greatest impact, and further decreased the work required to unwind the helix by ∼30 kJ mol^{-1} to 97.33 kJ mol^{-1}. Ten repeats were sufficient for the data to converge, as assessed by leave-one-out cross validation. Each data point was systematically excluded, and the mean was calculated. Omitting a data point does not result in the mean significantly changing (Fig. 5.10).

The trend observed agrees with the experimental data obtained using CD, in which the parent peptide was most helical, followed by the N-terminal oxetane-modified peptides, while the centrally-modified peptide was the least helical.[9,93] Using SMD to unwind peptides by pulling does not necessarily reflect a peptide's natural α-helical tendency—it is important to understand that in the simulations, the peptides are subjected to forces that do not occur under normal circumstances

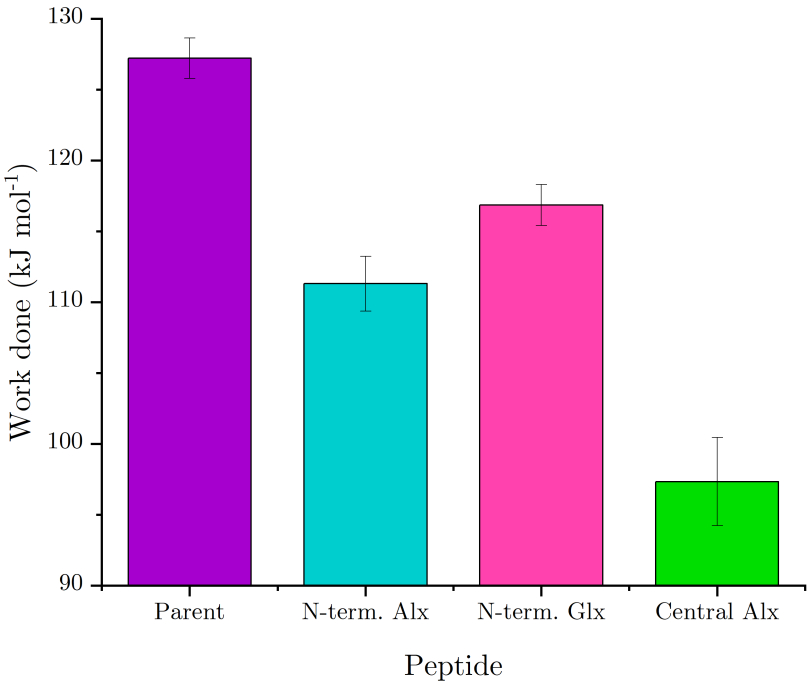


Figure 5.9: Work required to pull parent and oxetane-modified helices by 5 nm (20 ns of simulation). The parent peptide requires the most work to unwind, the N-term. Glx peptide is second, the N-term. Alx is third, and the central Alx peptide requires the least work to unwind. Error bars represent the standard error between 10 repeats, $p < 0.0001$ when compared using One-Way ANOVA.

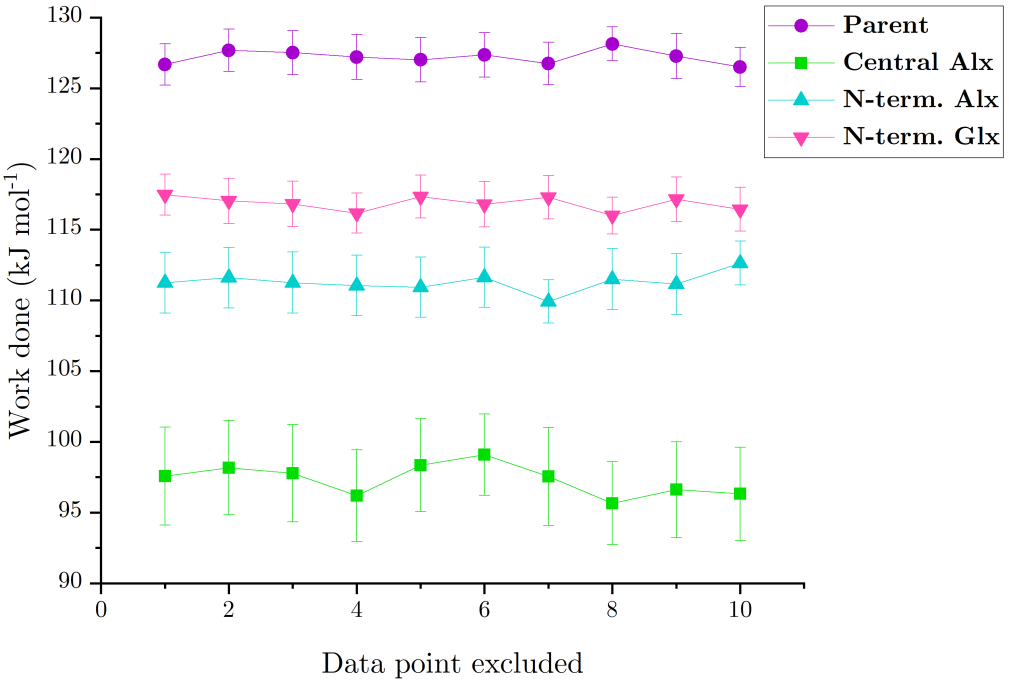


Figure 5.10: Leave-one-out cross-validation of simulation data for the ten repeats for each peptide. Systematic exclusion of a data point does not drastically affect the mean work done, and any changes remain within error.

and that the conditions used for the SMD simulations are entirely artificial. However, SMD can provide insight into regions of the peptide which unwind more readily and the amount of work required to do so. It is important to compare this to experimental data in order to validate the model, and in our work the trend matches closely. The SMD data indicates that the introduction of an oxetane into a highly helical region destabilizes the fold by up to 30 kJ mol^{-1}. For reference, the difference in energy between the native and unfolded state of a protein is typically in the region of 20–40 kJ mol^{-1}.[188] Therefore, oxetane is highly destabilising when inserted into a helical region, resulting in destabilisation several times higher than that caused by amino acid mutation,[189] or corresponding to the loss of multiple strong hydrogen bonds.[190] In comparison, thioamide introduction in the same position resulted in destabilisation of the fold by ∼7 kJ mol^{-1},[184] due to the longer C=S bond (1.56 Å)[191] and larger sulphur atom.

The trend observed when comparing glycine and alanine modification at position 3 is expected, as under normal conditions glycine is considered disruptive in internal helical positions due to its high conformational flexibility, and alanine is considered stabilising,[192] so oxetane modification alters these properties. It may be the case that glycine's flexibility and lack of bulky sidechain allow it to slightly compensate for the oxetane behaving as a conformational lock. However any compensation effect is minor, as ultimately the differences in helicity between N-term. Alx and N-term. Glx as measured by CD, and the differences in work required to unwind the two peptides as calculated using SMD, are small.

5.2.3.3 Oxetane Modification Alters the Structure and Hydrogen Bonding Pattern of Helical Peptides

The changes in the amount of work required to unwind the oxetane-modified peptides compared to the parent demonstrate a large destabilisation of the helical fold by introduction of a single oxetane modification. To explore the molecular determinants for this instability, the atomistic details of the simulation were probed. During the first 5 ns of the SMD simulations, the helices relaxed away from their ideal helical structure starting configurations, although the unmodified peptide remained highly helical.

Oxetane modification has a clear impact on the structure and hydrogen-bonding pattern of the helical peptides. Representative snapshots were prepared after 5 ns of simulation time (Fig. 5.11). This time point was selected as the position restraints have been removed and the peptide has had time to be released from the ideal helix starting configuration, but 90–100% of the starting hydrogen bonds remain intact, suggesting the pulling is not yet affecting the structures of the peptide.

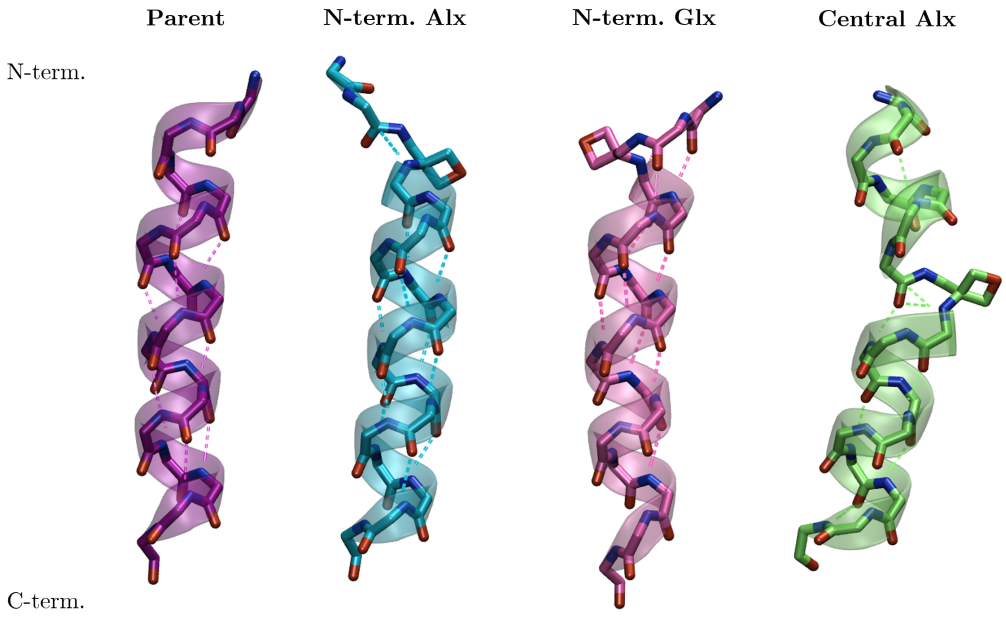


Figure 5.11: Snapshots of representative helical peptides taken at 5 ns, close to the start of the simulation. Helical structures for parent (purple), N-term. Alx (cyan), N-term. Glx (pink) and central Alx (green) show clear kinking in proximity to the oxetane modification, as well as changes in the hydrogen bonding (represented with dashed lines) patterns about and downstream of the modification.

There is distortion of the helical axis near the oxetane modification in all three modified peptides, which appears to produce a kink in the helix and changes the hydrogen-bonding pattern. This kinking effect may reflect the oxetane acting as a β-turn-inducing element, as observed in short linear oxetane-modified peptides (described in section 4.2.1.2).[92] In the parent peptide, the i, i+4 hydrogen bonds characteristic of an α-helix are maintained, whilst in the oxetane-modified peptides these hydrogen bonds are lost in proximity to the modification, and additional i, i+2 hydrogen bonds are formed.

Modification of the peptide backbone with an oxetane heterocycle changes the order in which the residues of the peptide unwind. In a mechanism generally conserved between the 10 repeats, the parent peptide typically unwinds slightly from the N-term., followed by the major unwinding events occurring from the C-term., and the central region of the peptide unwinds last (Fig. 5.12a). However, oxetane modification at residue 8 (Fig. 5.12d) promotes unwinding of the central region of the peptide, and the presence of the modification at residue 3 promotes unwinding at the N-term., in proximity to the modification (Fig. 5.12b and c).

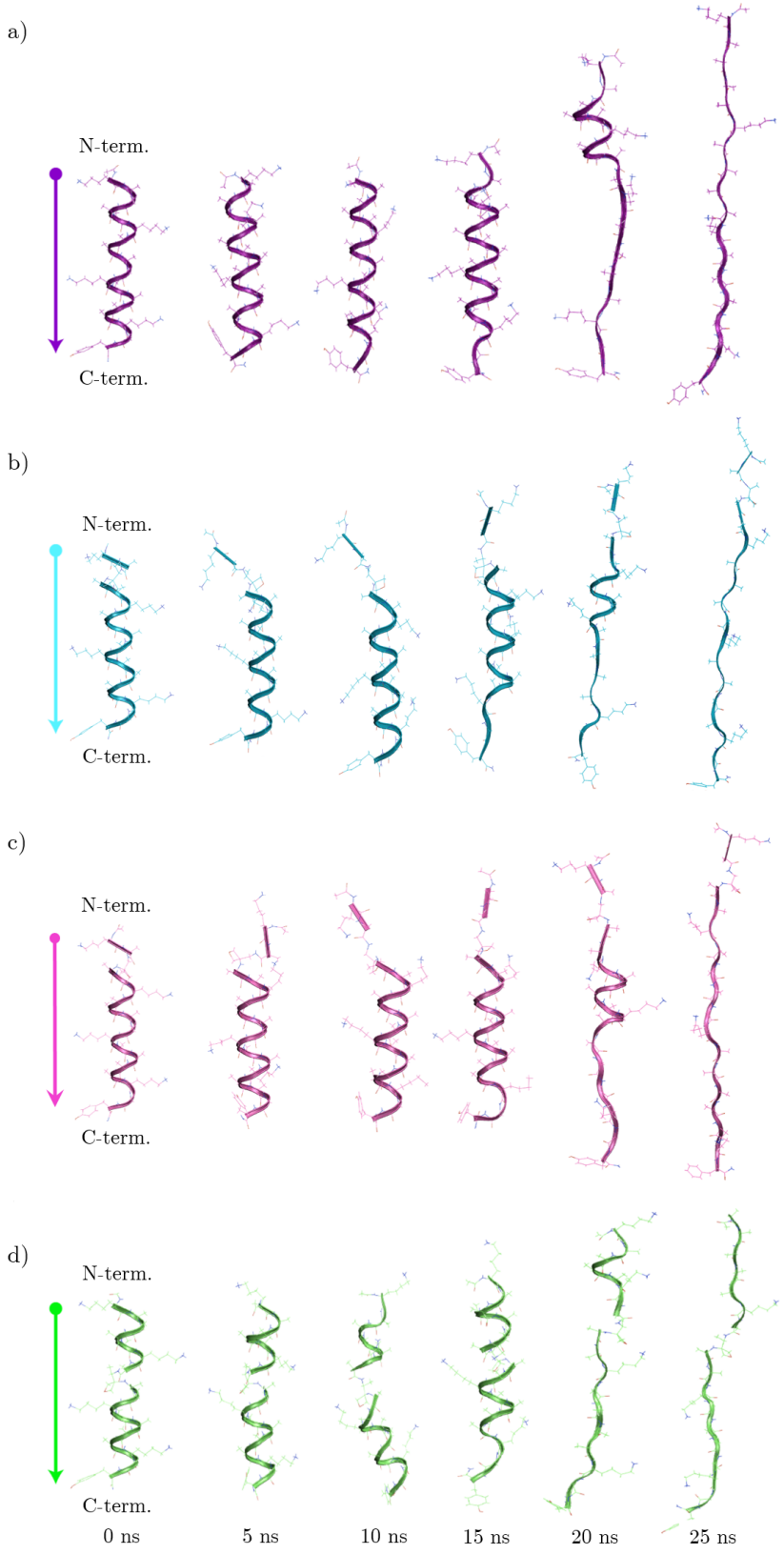


Figure 5.12: Snapshots of (a) parent, (b) N-term. Alx, (c) N-term. Glx, and (d) central Alx helical peptides taken at 5 ns intervals, highlighting the key unwinding events. In all cases, helicity is completely lost after 25 ns of simulation. Each peptide was anchored at by the Cα atom of Lys1, and pulled by the Cα atom of Tyr18.

5.2.3.4 Dihedral Angle Analysis Reveals Changes 2–3 Residues in Either Direction From the Modification

One of the characteristic properties of an α-helix is that all the amino acids have negative ϕ and ψ angles, typically around -60° and -40° respectively. Ramachandran plots were prepared for 2–5 ns of the trajectory. The initial 2 ns was omitted to allow the peptides to relax from their position-restrained ideal helix starting configurations, and, as previously discussed, at 5 ns the structures of the peptides are not yet affected by the pulling. The plots show the differences between the parent and N-terminally-modified peptides (Fig. 5.13), and between the parent and central Alx peptides (Fig. 5.14). While residues 3 to 11 of the parent correspond to the characteristic conformational space of an α-helix, in the presence of the oxetane modification there is a clear distortion in dihedral angles that spreads 2–3 residues in either direction from the modification. This agrees with the previously observed NMR data, which suggests that only one of Lys 1, 6 or 11 is in a helical environment. Disrupting helical content by 3 residues in either direction from the modification at position 3 would prevent Lys 1 and 6 from being in a helical environment.

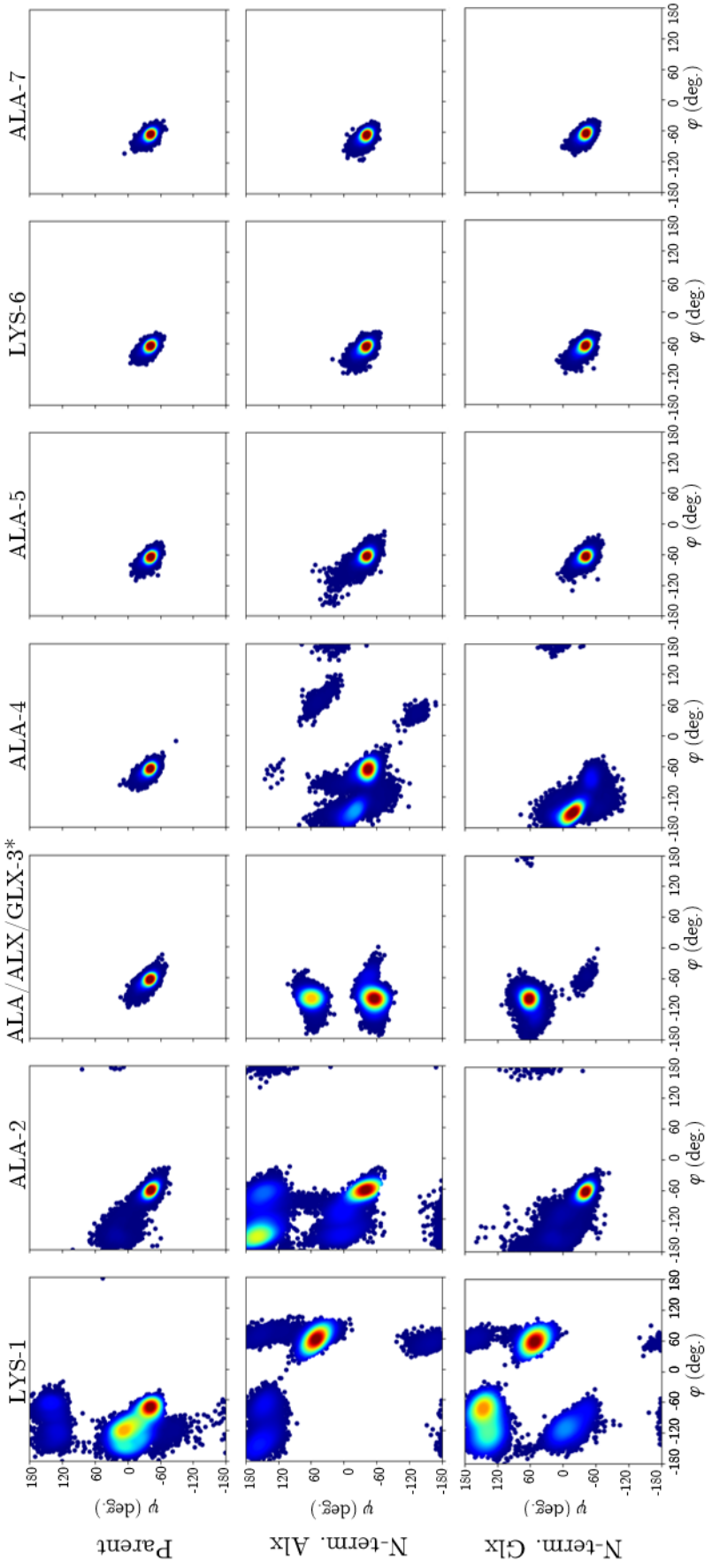


Figure 5.13: Ramachandran plots for residues 1–7 of parent, N-term. Alx and N-term. Glx peptides. Disruption from the oxetane modification at residue 3 extends in either direction by 2–3 residues.

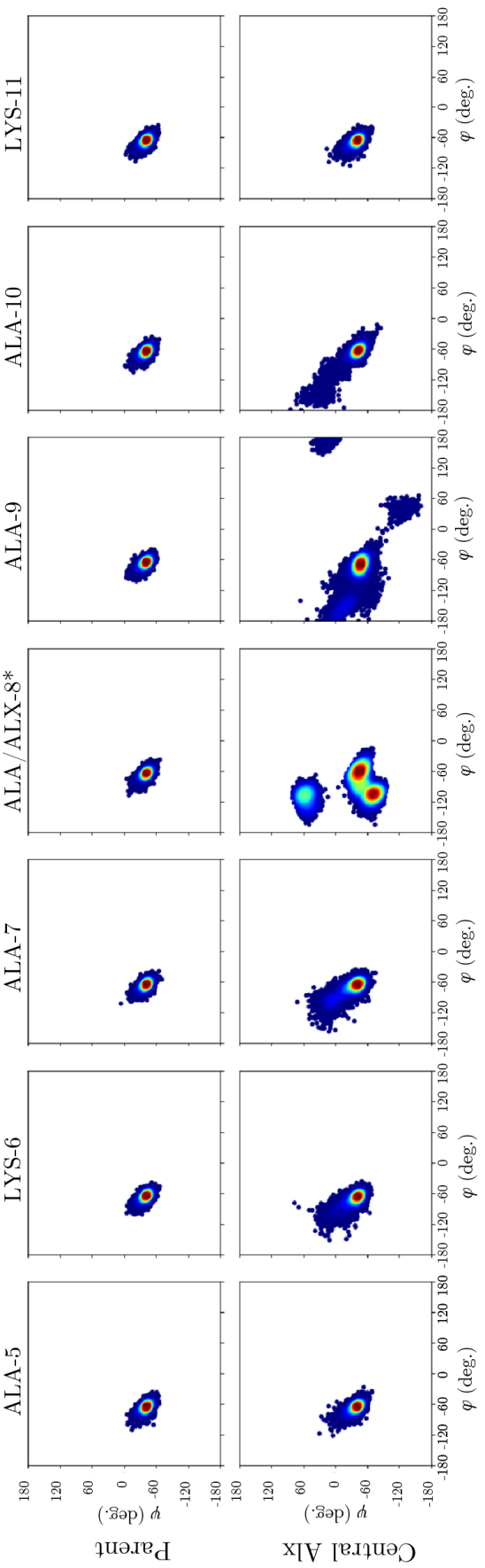


Figure 5.14: Ramachandran plots for residues 5–11 of parent and central Alx peptides. Disruption from the oxetane modification at residue 8 extends in either direction by 2–3 residues.

5.2.3.5 Comparison to Short Linear Peptides

The sampling of ϕ/ψ space in the Glx residue of the N-term. Glx peptide was compared to that in the short linear pentapeptide $LAG_{ox}AY$-OMe (described in Chapter 4 and previously by us[92]). Although it is difficult to directly compare the dihedral angles observed in the oxetane-modified helical peptides to those in $LAG_{ox}AY$-OMe due to differences in solvent, methodology and starting structures, there are some similarities between the Ramachandran plots (Fig. 5.15). Introduction of an oxetane modification in all cases results in a splitting of the ϕ/ψ space sampled, with two populations appearing: one with a negative ϕ/negative ψ, and the second with a ϕ of -120 to -180° and a positive ψ angle. Interestingly, the sampling of the ϕ/ψ space in the Glx residue of $LAG_{ox}AY$-OMe appears more similar to that of the Alx in the N-term. Alx peptide than to the Glx in the N-term. Glx peptide, however these distinctions may be arising due to fundamental differences in the systems as previously described.

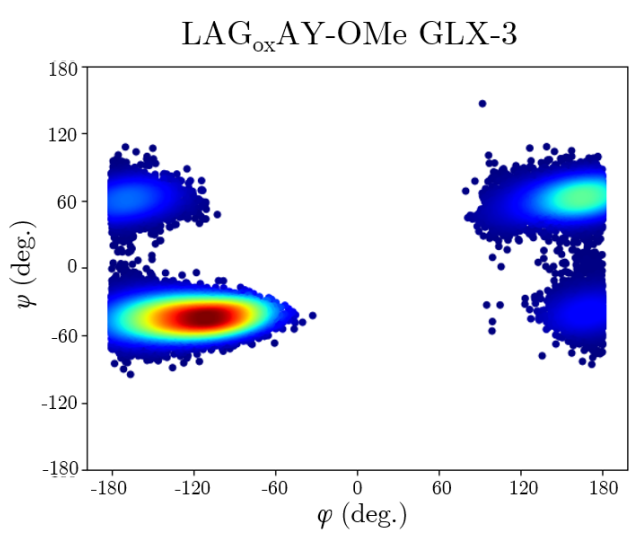


Figure 5.15: Ramachandran plot for Glx residue in short linear peptide, $LAG_{ox}AY$-OMe.

5.2.3.6 Hydrogen Bonding Analysis is Consistent with Experimental Data and Key Unwinding Events

The number of hydrogen bonds per residue over time were plotted for 20 ns of trajectory for all ten repeats. Representative plots are shown in Fig. 5.16 (data for all repeats is included in Appendices).

In the parent peptide, most residues form on average 1 hydrogen bond per residue. At the start of the simulation, there is evidence of unwinding of the helix at the N-terminus, and as pulling proceeds the peptide predominantly unwinds

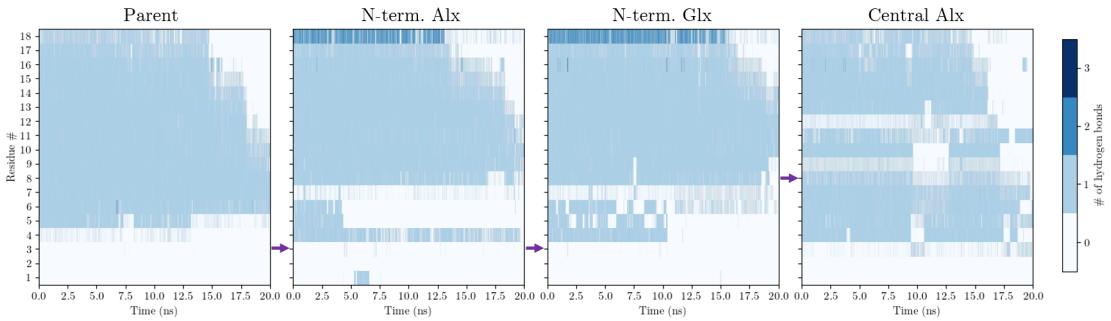


Figure 5.16: Representative plots indicating the number of hydrogen bonds per residue over 20 ns. Modified sites are indicated with an arrow. There is distortion up to four residues away from the site of modification.

from its C-terminus. This is in agreement with the key unwinding events observed during visual inspection of the trajectory (Fig. 5.12).

Oxetane modification at position 3 (both N-term. Alx and N-term. Glx) causes a disruption in the formation of hydrogen bonds four residues away (i.e. up to residue 7). In the case of the N-term. Glx peptide, the hydrogen bonds appear to be maintained for longer compared to the N-term. Alx, although there is a clear disruption in *i, i*+4 hydrogen bonds downstream of the modification (Fig. 5.16). For the N-term. Alx peptide, the disruption of the hydrogen bonding network is also visible in the snapshots taken at 5 ns (Fig. 5.11) and later in the trajectory, where the N-terminus of the peptide adopts a largely extended conformation (Fig. 5.12). As pulling proceeds, unwinding again occurs from the C-terminus. In the central Alx peptide (modification at residue 8), the hydrogen bonding patterns in the middle of the peptide are disrupted (Fig. 5.16), and there is a kinking effect in the helix, particularly obvious after 10 ns (Fig. 5.12). Like the parent peptide, the central Alx peptide tends to unwind slightly at the N-terminus, and then predominantly from the C-terminus, although it appears that oxetane modification promotes unwinding of the central region of the peptide, as this region unwinds more readily.

The disruption in hydrogen bonding described by both the number of hydrogen-bonds per residue and the key unwinding events are consistent with the experimental data[9,93] and suggest that the oxetane modification destabilises α-helicity by reducing the ability of the peptide to form the *i, i*+4 hydrogen bonds characteristic of an α-helix, as evidenced by their noticeable absence in proximity to the modification in the SMD structures.

5.3 Conclusions

The α-helix is one of the most common and most important structural motifs in biology, and it is necessary to characterise the impact of any peptide modification on secondary structure. CD has previously been used to show that the oxetane modification is highly disruptive to the helical fold and poorly tolerated in an α-helical structure,[9,93] despite the similarities in lone pair arrangement and hydrogen bonding capabilities between oxetanes and carbonyls.[53] This was confirmed by assessing chemical shift changes of the Hα atoms of the parent and N-term. Glx peptides. In order to gain insights into the molecular determinants for this instability, we used SMD simulations to artificially unwind the four peptides, and were able to replicate the observed experimental trend by calculating the amount of work required to unwind each peptide.

Analysis of the simulations suggests that the disruption to helicity is caused by changes in dihedral angles and hydrogen bonding patterns in proximity to the modification, due to the identity and size of the hydrogen bonding groups involved. While the α-helical structure is well-adapted to amide and carbonyl groups, backbone modification of an α-helix is not necessarily always disruptive—in fact, some backbone modifications have been shown to increase or induce helicity. For example, introduction of lactam bridges between *i, i*+4 amino acids has been shown to improve helix stability and bioactivity of human parathyroid hormone.[193] A more drastic backbone substitution, the replacing of a dipeptide in an α-helix with a leucine-derived 1,2,3-triazole ε2-amino acid, results in the modified peptides retaining much of the helical structure of the parent sequence.[194] Therefore, it would appear that as long as a modification is able to maintain *i, i*+4 hydrogen bonds, helix disruption is minimal, suggesting that helix modifications should be considered from a hydrogen bonding perspective, rather than how structurally conservative they may first appear.

Taken together, our data suggest that oxetane substitution disrupts helicity by changing the dihedral angles of the peptide backbone in the vicinity of the modified residue, such that they tend towards a β-turn and therefore disrupt the *i, i*+4 hydrogen bonds known to be vital for helix stability. This is consistent with work described in the preceding chapter, which suggests that in short linear peptides, the oxetane modification acts as a β-turn-inducer.[7,92] Other factors associated with the modified residue such as its increased molecular volume likely contribute to these distortions seen in the peptide backbone.

Oxetanes are therefore not a useful modification for stabilising helices due to their β-turn-inducing effect. This modification remains best-tolerated in peptides containing secondary structure involving turns, such as α-hairpins or β-turns,

where it has been demonstrated that the oxetane can enhance cyclisation efficiency and stabilise turn elements in peptides.[7,92] One promising direction is the use of oxetane modification to stabilise or mimic β-turn structures. Like α-helices, β-turn structural motifs are widespread in proteins. They have been implicated in molecular recognition and PPIs for proteins including GPCRs[195] and β-amyloid,[196] and thus are of considerable interest in the field of peptidomimetics for therapeutic use.

Chapter 6

Azetidine-Modification on Antimicrobial Peptides

6.1 Introduction

Antibiotic resistance is an increasingly significant threat to human health. Overuse and misuse of antibiotics in healthcare and agriculture has resulted in pathogenic resistance (Table 6.1, Fig. 6.1), and bacteria resistant to one treatment must be treated with another antibiotic, resulting in the emergence of multi-drug resistant bacteria. Of particular concern is the prevalence of multi-drug resistance in ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species), which are the leading cause of hospital-derived (nosocomial) infections.[78] A systematic review of antibiotic resistance showed that ESKAPE bacteria are associated with the highest risk of mortality, and that drug-resistant ESKAPE pathogens result in significantly increased healthcare costs.[197]

Table 6.1: Causes of antibiotic resistance. Adapted from World Health Organisation.[79]

Cause of antibiotic resistance
Overprescribing of antibiotics
Patients not finishing their treatments
Poor infection control in hospitals and clinics
Lack of hygiene and poor sanitation
Lack of new antibiotics being developed
Overuse of antibiotics in livestock and fish farming

From the late 1960s to the early 1980s, the pharmaceutical industry introduced many new antibiotics, but after that the antibiotic pipeline began to dry up and fewer drugs were introduced.[198] As a result, many decades after the start

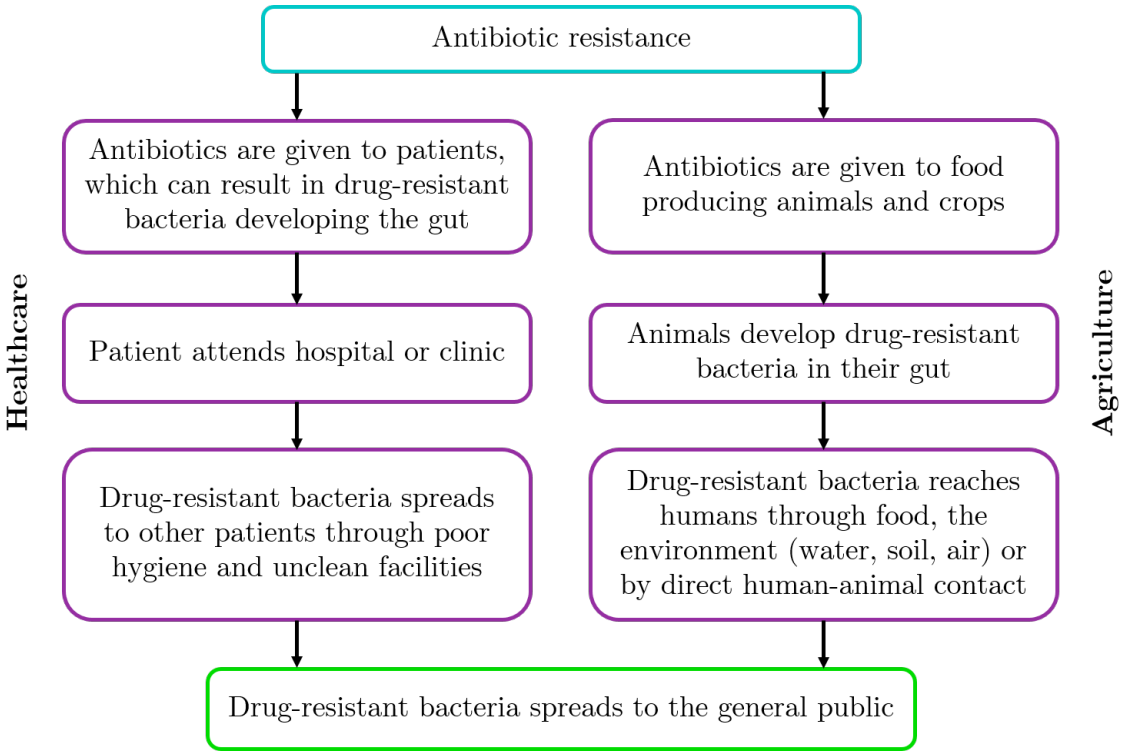


Figure 6.1: Mechanisms by which antibiotic resistance spreads. Adapted from World Health Organisation.[79]

of the modern era of antibiotics (following the discovery of penicillin in 1928 by Sir Alexander Fleming),[199] bacterial infections are once again a major threat to human health.[198] Unless action is taken, drug-resistant diseases are projected to cause 10 million deaths per year globally by 2050,[79] with increasing concerns about a 'post-antibiotic' era in which once-trivial infections can no longer be effectively treated.[80] Outside of healthcare, antibiotic resistance threatens four of the United Nations General Assembly's Sustainable Development Goals: by pushing an additional 24 million people into extreme poverty by 2030; by affecting livestock production and resulting in poor nutrition; by causing economic losses as large as the 2008–2009 financial crisis, amounting to $3.4 trillion by 2030; and by widening economic inequality between countries.[200]

There is an urgent need for the development of novel antimicrobials. Since their discovery in the 1980s, antimicrobial peptides (AMPs) represent a promising alternative to today's antibiotics.[85] AMPs, also called host defence peptides (HDPs), are part of the innate immunity of most living organisms.[80] They can exhibit a broad spectrum of activity against pathogenic bacteria and fungi, and can often demonstrate activity against viruses, parasites and cancerous cells.[85]

Importantly, AMPs have the potential to overcome bacterial resistance.[82] Many AMPs are cationic, such as polymyxin B, an FDA-approved antibiotic derived from *Bacillus polymyxa*. Cationic residues are thought to mediate the initial electro-

static attraction to the negatively charged bacterial cell membranes,[87] and arginine appears to be particularly favourable due to its ability to form hydrogen bonds when in a stacked cation-π arrangement with tryptophan.[88] These cation-π interactions effectively shield the Arg from the hydrophobic lipid bilayer, making entry more energetically favourable.[86] Tryptophan residues are also very common in AMPs, as Trp has a preference for the interfacial region of lipid bilayers, anchoring the peptide to the bilayer surface and allowing for prolonged association with the membrane.[87] By inserting themselves into the membrane, AMPs can form pores via the toroidal pore or barrel-stave pore models, or can disrupt the membrane through a detergent-like action via the carpet model (see section 1.6, Fig. 1.17), amongst other mechanisms of action. Although some AMPs do target intracellular processes such as DNA and protein synthesis,[85] it is their interactions with the membrane which prove most challenging for bacteria to evade via common resistance mechanisms (i.e. mutation).

Like many other peptide-based drugs, AMPs have some limitations, including susceptibility to proteases. Cyclisation and the use of peptidomimetics has previously been used to avoid degradation of AMPs,[78,80] and in many cases has been shown to enhance antimicrobial activity.[89–91] Compared to their linear counterparts, cyclic peptides demonstrate enhanced stability, low toxicity and enhanced binding properties, making cyclic AMPs attractive potential therapeutics and a growing area of interest.[89] Oxetane- or azetidine-modification can assist in the otherwise challenging preparation of cyclic peptides. However, oxetane is prone to ring-opening in the acidic conditions required to remove arginine sidechain protecting groups.[41] Since the vast majority of AMPs have at least one arginine residue, the conditions required for the preparation of oxetane-modified AMPs can be unfavourable. The azetidine heterocycle is much more stable under acidic conditions, and azetidine modification is expected to behave in the same turn-inducing manner as oxetane modification (see section 4.2.3), allowing for the preparation of short azetidine-modified linear and cyclic AMPs. Therefore, we aimed to characterise the impact of azetidine modification and cyclisation on the antimicrobial activity of a sequence with previously described activity.

In this chapter, peptide sequences reported in the literature[150,151] to display potent activity against methicillin-resistant *S. aureus* (MRSA) were screened to confirm activity. MRSA is an ESKAPE pathogen and a common cause of nosocomial infection. In our hands, our first selected target (bacaucin-1) and several derivatives of this peptide yielded no observable activity, which is at odds with previously published results[150,151] and was unexplainable from biophysical characterisation data. Therefore, five additional sequences[152–156] with published activities against *E. coli* or *S. aureus* were screened. Of these, one promising target was se-

lected for modification. The impact of cyclisation and azetidine modification was explored for this peptide sequence, and microbiological assays revealed that azetidine modification increases antimicrobial activity, without increasing haemolytic activity.

6.2 Results and Discussion

6.2.1 Screening of Bacaucin-1 Derivatives

The naturally-occurring cyclic lipopeptide bacaucin, derived from *Bacillus subtilis*, was reported to demonstrate promising broad spectrum activity against Gram-positive bacteria.[151] However, this peptide was also cytotoxic against mammalian cells, likely due to the hydrophobic fatty acids, which have previously been reported to result in non-specific cytotoxicity.[201] Linearisation and removal of the fatty acid moiety of bacaucin resulted in the peptide bacaucin-1, which was reported to increase specificity against *S. aureus* and MRSA,[151] and further work involved the use of alanine scanning to probe how each residue affected antimicrobial activity. One derivative, bacaucin-1a, demonstrated increased activity compared to bacaucin-1.[150]

Initially, we screened bacaucin-1, bacaucin-1a and two additional derivatives (Table 6.2) under the conditions described by Zhu and colleagues.[150,151] These derivatives were prepared with additional alanine or glycine residues at position 1 or 4 (Table 6.2), to make incorporation of the azetidine modification at different sites possible. It was unclear how these amino acid substitutions would affect the antimicrobial activity, but the original sequences of bacaucin-1 and bacaucin-1a were expected to demonstrate good activity (minimum inhibitory concentration (MIC) values of 2–4 μg/ml) against *S. aureus* ATCC 29213 and MRSA, amongst other Gram-positive organisms. They were not expected to exhibit good activity (MIC value of $>$ 128 μg/ml) against Gram-negative strains such as *E. coli*. However, in our hands, we observed no antimicrobial activity for any of the peptides against *E. coli* ATCC 25922 or *S. aureus* ATCC 29213 in cation-adjusted Mueller-Hinton broth (CaMHB). Therefore, activity was assessed using a variety of conditions, Table 6.3.

Initially, the assay was repeated using different media, to partially starve the cells, which would encourage the bacteria to take up any compounds in the media. This resulted in no activity, so peptides were solubilised in dimethyl sulphoxide (DMSO), a solvent known to increase membrane permeability[202] so that AMPs are better able to act at the bacterial membrane. However, even upon use of DMSO, no activity was observed. Similarly, addition of 0.1% TWEEN to help permeabilise

Table 6.2: Bacaucin-1 derivatives screened to assess biological activity, and their expected MIC values against *S. aureus* ATCC 29213 in CaMHB.

Peptide	Notes	Expected MIC (μg/ml)
ELLSRVD	Bacaucin-1, Liu et al.[151]	4
ALLSRVD	Bacaucin-1a, Liu et al.[150]	2
ALLARVD	No published activity	N/A
GLLSRVD	No published activity	N/A

Table 6.3: Conditions screened to assess antimicrobial activity of bacaucin-1 and bacaucin-1a. Under no conditions were MICs observed.

Condition	Notes
M9 minimal media (supplemented with maltose)	Semi-starvation, to force bacteria to take up compounds in media
$\frac{1}{2}$ strength peptone water (supplemented with lactose)	Semi-starvation, to force bacteria to take up compounds in media
Peptides dissolved in DMSO	Increases membrane permeability
TWEEN added to well	Detergent helps solubilise cell membrane
ΔTolC *E.coli*	Bacteria is missing an outer membrane channel protein and has a 'leaky' cell membrane

the bacterial membrane did not result in any observable activity. Finally, although the published sequences did not exhibit activity against Gram-negative bacteria, we also screened the peptides against a TolC-deficient strain of *E. coli*. TolC is an outer membrane efflux protein important in the export of small molecules across the outer membrane of Gram-negative bacteria,[203] so bacteria lacking this protein have a 'leaky' outer membrane. Along with partial starvation conditions, the lack of a TolC protein makes *E. coli* very vulnerable to antimicrobial compounds. Even in these extreme conditions, we did not observe any reduction in growth compared to untreated controls.

Low AMP concentration, resulting in a lack of activity, can be caused by poor peptide solubility. In order to confirm that the lack of activity observed was not due to low peptide concentration in our solutions, we used 1D proton NMR experiments, comparing the peak volume to those in 10 mM L-arginine, which was used as an external standard (see Appendices). This revealed the concentration of the peptide stock solution to be lower than expected at ∼3 mM; however this concentration should still be sufficient to observe antimicrobial activity. We also assessed the amino acid sequence using 2D ^{1}H-^{1}H TOCSY and ROESY spectra. Due to the size of the molecule, no NOEs were observed, but ROEs enabled us to perform a backbone walk and confirm that the peptide sequence was in the correct order. We then contacted Dr Zhu, who in a private correspondence to us confirmed that the authors of the previous work[150,151] were unable to replicate the

published activity of bacaucin-1 and bacaucin-1a in separate synthetic lots of the peptide.

6.2.2 Screening of a Panel of Short Antimicrobial Peptides

A panel of five additional peptides, reported to have antimicrobial activity, were selected following literature searches to screen against *E. coli* ATCC 25922 and *S. aureus* ATCC 29213 (Table 6.4). The selection criteria were that each peptide must be no more than seven amino acids in length (as sequences longer than this are substantially easier to cyclise using conventional methods[49,50]) and must contain an alanine or a glycine residue (to allow for azetidine modification). Three of the peptides were expected to exhibit activity against Gram-negative bacteria, and all five were expected to exhibit activity against Gram-positive bacteria. For two of the peptides (RRW and GLL), the mechanism of action has not been previously described,[152,153] while KNK, ALL and KVF were all expected to act at the bacterial membrane and induce lysis.[154–156]

Table 6.4: Subsequent panel of five AMPs, with notes on published activity and references. Abbreviations in brackets are used throughout Chapter 6.

Peptide sequence	Activity against	Reference
Ac-RRWWCA-NH_2 (RRW)	*S. aureus* *E. coli* *S. sanguinis* *P. aeruginosa* *C. albicans*	Blondelle et al.[152]
GLLKRIK-NH_2 (GLL)	*S. aureus* ATCC 25923 *E. coli* ATCC 25922	Ifrah et al.[153]
KNKGWWW (KNK)	*S. aureus* ATCC 29213 *E. coli* ATCC 25922	Pasupuleti et al.[154]
ALLRL (ALL)	Gram+ bacteria	Otsuka et al.[155]
KVFLGLK (KVF)	*S. aureus* ATCC 25923 *S. typhimurium* *P. aeruginosa* *S. dysenteriae* *B. subtilis* *S. pneumoniae*	Xiao et al.[156]

However, none of the peptides screened demonstrated activity against *E. coli* ATCC 25922 and *S. aureus* ATCC 29213 in CaMHB. In order to assess whether the peptides would demonstrate activity under extreme conditions, the five peptides were screened against *E. coli* ΔTolC in minimal media. A reduction in growth was observed for three of the peptides which were expected to be active against

Gram-negative bacteria (RRW, GLL and KNK), with MICs of 16, 256 and 128 μg/ml respectively (Table 6.5).

Table 6.5: Minimum inhibitory concentration (MIC) values for the panel of five antimicrobial peptides (AMPs) against TolC-deficient *E. coli* in minimal media.

Peptide	MIC (μg/ml)
Ac-RRWWCA-NH_2 (RRW)	16
GLLKRIK-NH_2 (GLL)	256
KNKGWWW (KNK)	128
ALLRL (ALL)	> 256
KVFLGLK (KVF)	> 256

RRW showed signs of peptide aggregation at concentrations of > 64 μg/ml. As this peptide contained a cysteine residue, it seems likely that disulphide bond formation was occurring at high peptide concentrations. In their previous work with this peptide, Blondelle et al. added 1.25 mM dithiothreitol (DTT) to each well prior to addition of bacterial suspension in order to prevent disulphide bond formation,[152] therefore moving forward RRW was screened in the presence and absence of DTT.

Following observation of activity in *E. coli* ΔTolC, we then aimed to elicit a response to these peptides against more clinically relevant strains. We screened the peptides against *E. coli* ATCC 25922 and *S. aureus* ATCC 29213 in minimal media, and added a small amount of sterile DMSO ($< 2.5\%$ and $< 5\%$ for each strain respectively) to help AMP uptake by the cells. DMSO is toxic to *E. coli* ATCC 25922 and *S. aureus* ATCC 29213 at $\sim$6% and $\sim$12.5% respectively, so these lower concentrations help permeabilise the bacterial membrane without directly killing the cells. For all experiments using DMSO, an additional positive growth control was used to ensure that the DMSO was not affecting the bacterial growth.

Under these conditions, no growth was observed after 48 hours for *S. aureus*, even in the positive growth controls. Addition of a small amount of GC media, a nutrient-rich media used to cultivate *Neisseria gonorrhoeae*, added after 24 hours resulted in visible reduction of growth compared to controls for carbenicillin, RRW and RRW + 1 mM DTT (MICs of 8, 256 and 256 μg/ml respectively). No other peptides demonstrated activity under these conditions, and addition of GC media at the start of the assay resulted in no reduction of growth for any of the peptides. The control antibiotic carbenicillin did result in a reduction of growth with an MIC of 16 μg/ml.

However, these conditions appeared to be suitable for observing a reduction of growth after 18–20 hours in *E. coli* ATCC 25922. MICs were observed for carbenicillin and RRW (in the presence and absence of DTT). DTT drastically

increased the activity of RRW, to the extent that no growth was observed even at peptide concentrations of less than 0.5 μg/ml. In order to confirm that DTT was not inhibiting growth alone, the concentration of DTT was reduced down to 0.25 mM (substantially lower than the 1.25 mM used by Blondelle et al.[152]), and it was also used without the presence of any AMP as a growth control. Additionally, the plate was incubated at 37 °C for a further 24 hours, and the wells containing RRW + DTT exhibited normal growth.

Therefore, for the three sequences which were expected to show activity against Gram-negative bacteria (RRW, GLL and KNK), some activity was observed, albeit at much higher concentrations and under more extreme conditions than expected. No activity was observed under any of the conditions tested for the other two peptides (ALL and KVF), which were expected to demonstrate some antimicrobial activity against Gram-positive bacteria. One peptide (RRW) demonstrated activity against both *E. coli* and *S. aureus* under specific conditions. For this peptide, activity was previously reported as an IC50 value (35–40 and 21–23 μg/ml for *S. aureus* ATCC 29213 and *E. coli* ATCC 25922 respectively). IC50 values cannot be directly compared to MIC values, as IC50 reports the concentration required to inhibit a biological process by 50%, while MIC reports the lowest concentration required to inhibit the visible growth of bacteria.

6.2.3 Biological Activity of RRW Derivatives

As RRW, a hexapeptide derived from synthetic combinatorial libraries,[152] was the most promising candidate, a series of derivatives were prepared (Table 6.6). These included a cyclic variant of the original sequence (for which activity has not been previously reported), and cyclic and linear variants in which the cysteine residue was replaced with a glycine. Since disulphide bonds appear to have a negative effect on the activity of the parent sequence (as addition of DTT improves activity), we hypothesised that changing this residue to a glycine would not have a drastic effect on the antimicrobial activity, and additionally would provide a site for azetidine modification. Two azetidine-modified variants were also prepared (one cyclic and one linear). The linear azetidine-modified variant, RWWG$_{az}$AR, does not have the exact same sequence from N- to C-terminus as the parent peptide RRW due to challenges in incorporating an azetidine-modified residue near a terminus (azetidine-facilitated cyclisation has been shown to be most effective when the modification is in the centre of the sequence[41]). For this reason, it was not possible to directly compare behaviour of RWWG$_{az}$AR to the original sequence of Ac-RRWWCA-NH$_2$, but it can be compared to the cyclic peptide cRRWWG$_{az}$A.

Table 6.6: Derivatives of RRW used to assess the effect of azetidine modification and/or cyclisation.

Peptide	Notes
Ac-RRWWCA-NH_2	Original sequence
cRRWWCA	Cyclised variant of original sequence
Ac-RRWWGA-NH_2	Cysteine replaced with glycine
cRRWWGA	Cyclised variant, contains glycine
cRRWWG_{az}A	Cyclised azetidine-modified peptide
RWWG_{az}AR	Azetidine-modified variant, sequence in different order due to challenges with preparation

6.2.3.1 Cyclisation Affects Antimicrobial Activity

Insertion of molecules into the lipid bilayer requires overcoming entropic and enthalpic energy barriers associated with a loss in conformational freedom. Therefore cyclisation can enhance the ability for AMPs to act at the membrane by restricting the flexibility of the molecule and reducing the depth of the entopic barrier.[204] In the past, cyclisation has been shown to increase antimicrobial activity in several cases. Dathe and colleagues showed that the sequence Ac-RRWWRF-NH_2 demonstrates poor activity against *E. coli*, while the cyclic analogue cRRWWRF is much more potent.[91] Mika and co-workers showed using *in vitro* assays and MD simulations that the cyclic peptide BPC194 more readily disrupts membranes than its linear counterparts, and demonstrates higher antimicrobial activity *in vivo*.[90] Similarly, MD simulation work by Cirac et al. suggests that cyclic peptides may embed into the membrane bilayer more deeply, forming toroidal pores whilst their linear counterparts remain at the surface.[205]

Here, we tested the effect of macrocyclisation on the biological activity of Ac-RRWWCA-NH_2. We compared the inhibition of growth of *E. coli* ATCC 25922 following treatment with Ac-RRWWCA-NH_2 and cRRWWCA in the presence and absence of DTT (Fig. 6.2; Table 6.7). In the absence of DTT, we observed a two-fold increase in the MIC of the linear peptide Ac-RRWWCA-NH_2 (MIC 128 μg/ml) compared to the cyclic variant cRRWWCA (64 μg/ml), which is in agreement with predictions based on the studies above.

However, in the presence of DTT this trend was lost. Addition of DTT to the linear sequence drastically increases activity (MIC of < 1 μg/ml compared to 128 μg/ml without DTT). Interestingly, the effect of DTT on the cyclic peptide was not consistent between repeats. Initially, addition of DTT increased the MIC (i.e. decreased activity) from 64 μg/ml to 256 μg/ml. On subsequent repeats, there was no visible bacterial growth after 24 hours in wells containing DTT at peptide concentrations of > 64 μg/ml and < 32 μg/ml, although there was high well turbid-

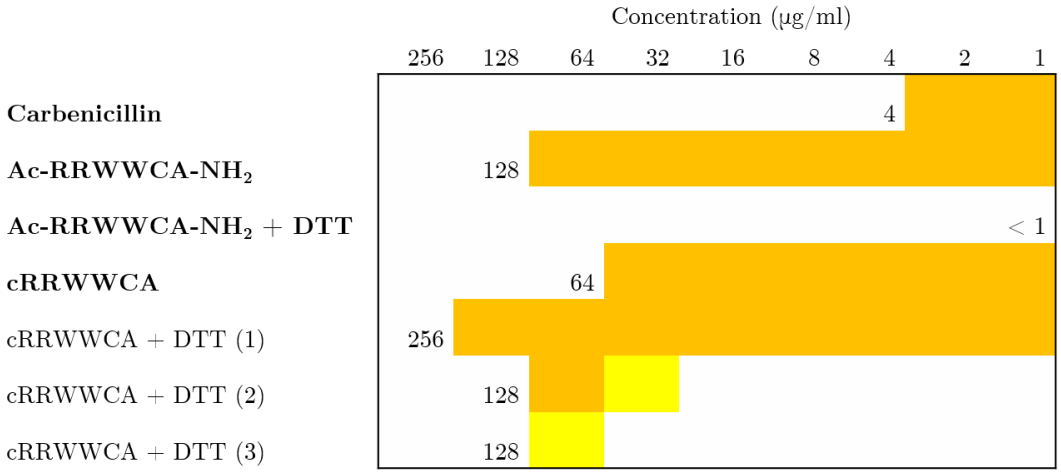


Figure 6.2: Schematic representation of bacterial growth after 18–20 hours at 37 °C. Control wells are omitted for clarity. Orange cells indicate visible bacterial growth, yellow represents medium turbidity (possible bacterial growth) and white represents no growth. MIC values are indicated with text. Bold indicates 3 separate biological repeats where the results are consistent across all repeats (cRRWWCA + DTT repeats are shown separately).

Table 6.7: Minimum inhibitory concentration (MIC) values for the RRW derivatives against *E. coli* ATCC 25922 in minimal media. Rows containing DTT do not have a real MIC value associated with them, (*) indicates partial turbidity in wells at a lower concentration. $n = 3$ biological repeats.

Compound	MIC (μg/ml)
Carbenicillin	4
Ac-RRWWCA-NH_2	128
Ac-RRWWCA-NH_2 + DTT	< 0.5
cRRWWCA	64
cRRWWCA + DTT	256–128*

ity (potentially indicating growth) at concentrations of ∼64 μg/ml. Comparison of preparative and analytical high performance liquid chromatography (HPLC) chromatograms indicated that this compound appeared to be changing over time, potentially through self-association. Samples were kept at -20 °C between use to minimise any degradation, but this was clearly not sufficient to prevent cRRWWCA changing in a way that affected its activity in the presence of DTT.

6.2.3.2 Azetidine Modification Increases Antimicrobial Activity

We then assessed the impact of substituting the cysteine with a glycine residue (Fig. 6.3; Table 6.8), expecting Ac-RRWWGA-NH_2 to retain much of the activity of the parent peptide, since the capability of the parent peptide to form disulphide bonds appeared to be detrimental to its activity. Surprisingly, activity was lost upon substitution of cysteine to glycine (MIC of $>$ 256 μg/ml, compared to 64 μg/ml for parent). Unlike the parent peptide RRW, where cyclisation altered the antimicrobial activity, cRRWWGA did not demonstrate improved activity compared to its linear counterpart. Therefore, it appears that the cysteine in Ac-RRWWCA-NH_2 is important for its antimicrobial activity. Previous work by Chen and co-workers showed that the presence of C-terminal cysteines can enhance antimicrobial activity.[206] Many antimicrobial peptides are rich in cysteine residues, although these sequences are typically much larger (40–100 amino acids in length).[207] In cysteine-rich peptides, disulphide bonds have been reported to promote antimicrobial activity and increase tolerance to high temperatures.[208]

Despite the lack of activity observed in the glycine-containing peptides, we assessed the effect of azetidine modification on these sequences (Fig. 6.3; Table 6.8). Intriguingly, azetidine modification of cRRWWGA (cRRWWG$_{az}$A) recovered observable antimicrobial activity against *E. coli* ATCC 25922 (MIC of 128 μg/ml), despite the unmodified sequence demonstrating no activity. This suggests that the polarity of the residue in this position of the chain is important for antimicrobial activity. Similarly, RWWG$_{az}$AR demonstrated improved activity, and is more potent than all other peptides screened in the absence of DTT, with an MIC of 32 μg/ml, although it is unclear to what extent this activity is due to the azetidine modification, as an unmodified variant of this sequence was not screened.

6.2.3.3 Azetidine-Modified Antimicrobial Peptides are Bacteriostatic

No data on bactericidal activity has previously been reported for any of the RRW derivatives (including the parent sequence). We performed a minimum bactericidal concentration (MBC) assay to assess the activity of the RRW derivatives against *E. coli* ATCC 25922 (Fig. 6.4; Table 6.9). This assay involves adding solutions

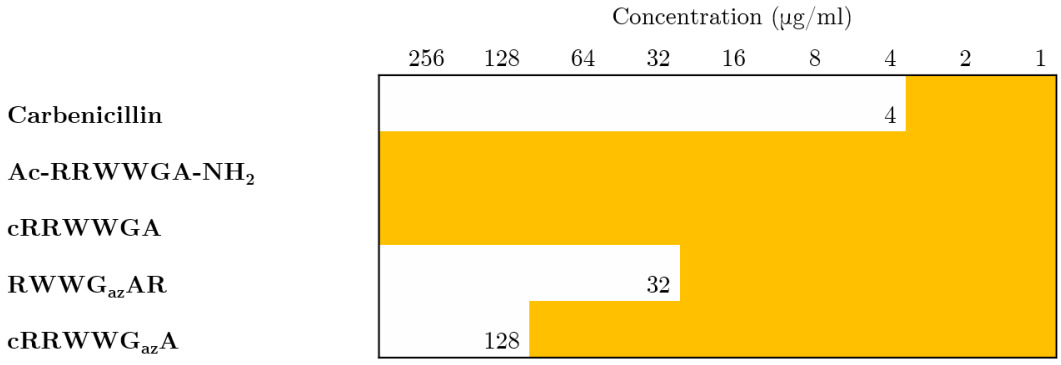


Figure 6.3: Schematic representation of bacterial growth after 18–20 hours at 37 °C. Control wells are omitted for clarity. Orange cells indicate visible bacterial growth and white represents no growth. MIC values are indicated with text. Bold indicates 3 separate biological repeats where the results are consistent across all repeats.

Table 6.8: Minimum inhibitory concentration (MIC) values for the glycine-substituted RRW derivatives against *E. coli* ATCC 25922 in minimal media.

Compound	MIC (μg/ml)
Carbenicillin	4
Ac-RRWWGA-NH_2	> 256
cRRWWGA	> 256
RWWG$_{az}$AR	32
cRRWWG$_{az}$A	128

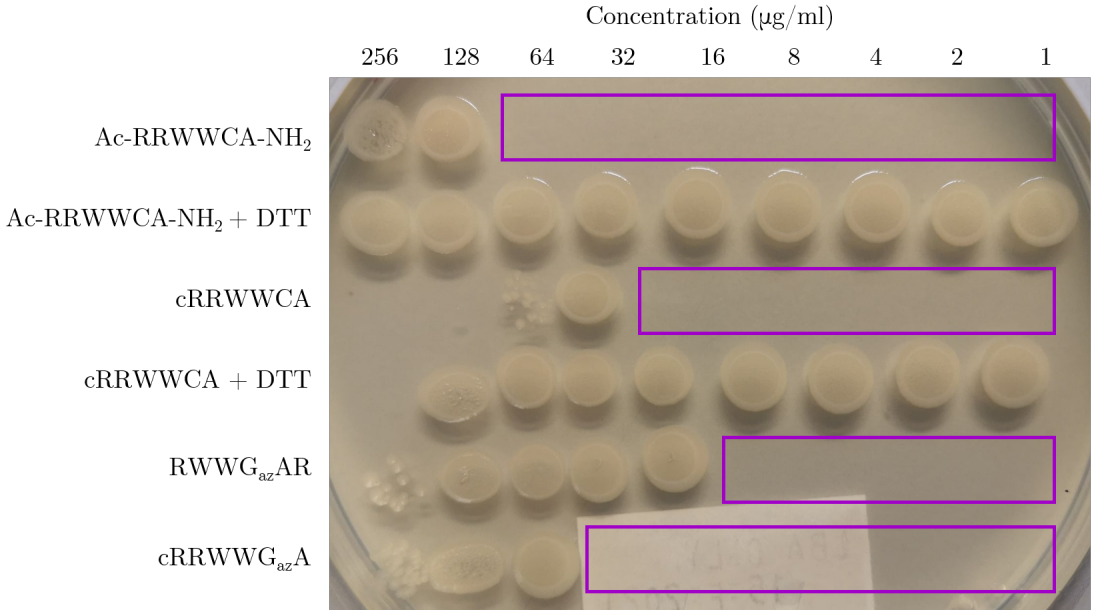


Figure 6.4: Minimum bactericidal concentration (MBC) assay. Solutions containing bacterial inoculum and compounds at concentrations equal to or higher than the MIC were taken from the MIC 96-well plate and added dropwise onto an agar plate. The MBC is equal to the lowest concentration where no growth is observed. Purple boxes indicate no solution was added (as these concentrations were lower than the MIC).

containing bacterial inoculum and peptides at concentrations equal to or higher than the MIC dropwise onto an agar plate and being left to grow for 24 hours at 37 °C. Where no growth occurs, this is because the bacteria are not viable following treatment with the compound (i.e. the antimicrobial compound has killed the bacteria). Therefore, this technique reports on the minimum concentration at which the compounds are bactericidal. This technique provides a different measure to an MIC assay, which reports the minimum concentration required to inhibit visible bacterial growth (without necessarily killing the cells).

Table 6.9: Minimum bactericidal concentration (MBC) values for the RRW derivatives against *E. coli* ATCC 25922.

Compound	MBC (μg/ml)
Ac-RRWWCA-NH_2	> 256
Ac-RRWWCA-NH_2 + DTT	> 256
cRRWWCA	128
cRRWWCA + DTT	256
RWWG$_{az}$AR	> 256
cRRWWG$_{az}$A	> 256

The parent sequence demonstrates poor bactericidal activity both with and without DTT (MBC of > 256 μg/ml), while the cyclic variant cRRWWCA has

improved bactericidal activity under both conditions (MBC values of 128 and 256 μg/ml in the absence and presence of DTT respectively). The peptide cRRWWCA would be considered bactericidal (i.e. kills bacteria), as the ratio of MBC to MIC is less than 4.[209] Both azetidine-modified peptides have low bactericidal activity (MBC of $>$ 256 μg/ml for both). Small molecule azetidine derivatives have previously been shown to exert promising antimicrobial activities,[210,211] but the mechanisms by which azetidine increases antimicrobial activity of a compound remains unclear. Based on the MIC and MBC assays, these azetidine-modified peptides appear to be bacteriostatic (i.e. prevent bacterial growth by keeping them in the stationary phase) rather than bactericidal (i.e. killing the bacteria). However, bactericidal agents are not necessarily superior to bacteriostatic agents. Potentially adverse clinical consequences can result from the rapid lytic action of bactericidal antimicrobials,[212,213] such as endotoxin surge in the cerebrospinal fluid of infants with Gram-negative meningitis following treatment.[214,215] Bacteriostatic agents are also able to inhibit protein synthesis in slow-growing bacteria not affected by bactericidal β-lactams. In fact, bacteriostatic antimicrobials (e.g. chloramphenicol, clindamycin and linezolid) have been successfully used to treat infections that are often considered to require bactericidal activity such as meningitis, endocarditis, and osteomyelitis.[209]

Although the mechanism of action for the parent sequence has not been elucidated, one might hypothesis that it would act at the bacterial membrane, considering the high content of Trp and Arg residues. These residues are vital for establishing strong interactions with the membrane.[86,87,152] However, targeting the membrane often results cell death, and the parent sequence demonstrates bacteriostatic rather than bactericidal activity, suggesting that it may have an intracellular target instead.

Although cysteine residues have been implicated in the targeting and uptake of cell-penetrating peptides,[216] cysteine-containing peptides have also been reported to target intracellular processes, such as inhibiting proteases.[217] Additionally, the parent peptide did not demonstrate any activity except against cells with weakened membranes (either against *E. coli* ΔTolC or in the presence of DMSO), suggesting that the membrane must be 'leaky' in order for this peptide to exert its activity inside the cell. Substituting the cysteine residue to an azetidine-modified glycine removes the capability to form disulphide bonds but keeps the residue relatively polar, although this does not enhance its bactericidal activity, suggesting it still acts on an intracellular target.

6.2.3.4 Azetidine Modification Does Not Increase Haemolytic Activity

In order for an antimicrobial compound to be of therapeutic use, it is important that it is specific to prokaryotic strains and does not result in cytotoxicity of mammalian cells. Toxicity of antimicrobial agents can be assessed through a haemolysis assay, in which the compounds of interest are incubated with mammalian erythrocytes. No haemolysis data had previously been described for any of the RRW derivatives (including the parent sequence). We performed a haemolysis assay to assess the lytic activity of the RRW derivatives on horse red blood cells. None of the peptides demonstrated haemolytic activity even at high concentrations after 2 hours. After 24 hours, four of the peptides demonstrated some haemolytic activity at high concentrations (Table 6.10).

Table 6.10: Minimum concentrations resulting in haemolysis of horse erythrocytes after 2 and 24 hours.

Compound	Minimum concentration of haemolysis (μg/ml)	
	2 hours	24 hours
Ac-RRWWCA-NH_2	> 1024	> 1024
Ac-RRWWCA-NH_2 + DTT	> 1024	> 1024
cRRWWCA	> 1024	512
cRRWWCA + DTT	> 1024	512
Ac-RRWWGA-NH_2	> 1024	> 1024
cRRWWGA	> 1024	1024
RWWG$_{az}$AR	> 1024	1024
cRRWWG$_{az}$A	> 1024	1024

The original sequence Ac-RRWWCA-NH_2 does not exhibit any lytic activity against horse red blood cells even after 24 hours, in both the presence and absence of DTT. The cyclised variant, cRRWWCA, is the most lytic after 24 hours out of all the peptides screened. This peptide is also the only one which could be considered bactericidal, suggesting that it is not highly specific in its cytotoxicity. Interestingly, treatment with the peptide cRRWWGA did result in some haemolysis after 24 hours at concentrations of 1024 μg/ml, despite this peptide not demonstrating any antimicrobial activity. Similarly, both the azetidine-modified variants did cause haemolysis after 24 hours at high concentrations. Despite this, the peptide RWWG$_{az}$AR is of potential therapeutic use, as there is no lysis within 2 hours, and after 24 hours the concentration which results in haemolysis is still more than 20 times that required to inhibit growth of *E. coli* 25922. However, its ability as a potential therapeutic is limited by its relatively poor antimicrobial activity, which is lower than that of many other antimicrobial peptides. For example, variants of Lynronne-1, a 19-residue amphipathic helix first isolated from

the rumen microbiome,[83] demonstrate potent antimicrobial activity (in the region of 8–1 μg/ml against ESKAPE pathogens *A. baumannii* and *S. aureus*).[218]

Taken together, the data suggest that azetidine modification increases antimicrobial activity against *E. coli* 25922 without resulting in increased toxicity. These microbiological assays could be complemented with microscopy experiments or membrane lysis assays to explore the mechanism by which azetidine modification exerts this effect.

6.3 Conclusions

The antibiotic resistance crisis is a significant threat to human health, with inappropriate use of antibiotics in healthcare and agriculture resulting in resistance to nearly all antibiotics that have been developed. AMPs represent a promising alternative to today's antibiotics,[85] particularly as they have the potential to overcome bacterial resistance.[82] Like many other peptide-based drugs, linear unmodified AMPs have some limitations, such as poor resistance to proteases, and as a result peptide cyclisation and backbone modification have previously been used to overcome the disadvantageous pharmacokinetic properties of AMPs.[78,80] Therefore, in order to explore the effect of cyclisation and/or azetidine modification on a peptide with documented antimicrobial activity, we used microbiological assays to screen a series of derivatives, and found that azetidine modification increases antimicrobial activity without increasing toxicity.

In order to maximise the impact of the azetidine modification, an ideal model system would contain less than seven amino acids (as sequences longer than this are substantially easier to cyclise[49,50]), and would contain a glycine or alanine residue to allow for modification. Initially, we selected bacaucin-1 and bacaucin-1a as these peptides were previously reported to demonstrate excellent antimicrobial activity against MRSA,[150,151] a common cause of nosocomial infection. However, we were unable to observe any antimicrobial activity for these peptides across a variety of conditions and bacterial strains. Instead, we screened a further five AMPs, and were able to elicit activity for three under extreme conditions. The most promising candidate, RRW, was selected for further screening, and a series of derivatives were prepared.

A cyclic derivative of RRW demonstrated increased activity compared to its linear counterpart, in line with previous observations that cyclisation often increases the antimicrobial activity of AMPs.[80,90,91,159] Substitution of cysteine to glycine resulted in the complete loss of activity for both linear and cyclic peptides, but this activity was restored following azetidine modification on the glycine residue. The

azetidine-modified AMPs demonstrated bacteriostatic activity rather than bactericidal activity, which is not necessary disadvantageous, as many diseases can be successfully treated with bacteriostatic agents.[209] Importantly, although azetidine modification restored antimicrobial activity compared to cRRWWGA (which demonstrated a complete loss of activity following substitution from cysteine to glycine), it did not result in increased haemolysis, which is important for a peptide with therapeutic implications.

Considering the high content of Trp and Arg residues in these sequences, we might expect the peptides to act at the bacterial membrane, as these residues are important in establishing strong initial interactions between the peptide and the negatively charged phospholipids of the bacterial membrane.[86,87,152] However, AMPs that are membrane-active are often bactericidal, as lysis of the membrane results in cell death. Traditionally AMPs are studied in the context of their ability to kill bacteria by disrupting the membrane,[219] but in reality AMPs are a widely variable group of peptides with a variety of complex mechanisms of action. In fact, AMPs have been implicated in the inhibition of DNA, RNA and protein synthesis, activation of autolysins (bacterial enzymes which break down peptidoglycan), and inhibition of enzymatic activity, amongst others cellular processes.[220]

Therefore, it is possible that these peptides do not act at the membrane and instead are involved in inhibiting bacterial growth by altering intracellular processes. This may explain why DMSO was necessary to elicit a response for the original sequence Ac-RRWWCA-NH_2, as the peptide was unable to exert its antimicrobial effect without increasing the permeability of the bacterial membrane. Of course, this has real-world implications and makes this peptide unlikely to be suitable as a therapeutic. Mechanisms of action of AMPs which do not act at the membrane can be explored using a variety of techniques, including gel electrophoresis to assess any interaction with DNA or RNA or co-precipitation assays to monitor interaction with nucleotides and proteins.[219]

Together, the work presented in this chapter shows that it is possible to produce azetidine-modified peptides that demonstrate antimicrobial activity, even if the activity is relatively poor and requires non-traditional conditions. Further work to optimise the conditions required to elicit an antimicrobial response, as well as studies to explore the mechanism of action of these peptides, could help explain how azetidine modification increases antimicrobial activity (without increasing toxicity), and this may be beneficial in the future design of modified antimicrobial peptides for therapeutic use.

Chapter 7

Conclusions and Further Work

7.1 Recap of Project Background and Aims

From the 1960s to the turn of the millennium peptides were often considered the drugs of the future.[24] Peptide-based drugs have the potential to combine the advantages of traditional small molecule drugs (such as oral bioavailability) with those of larger biologics (such as low toxicity).[16] However, naturally-occurring peptides generally have some challenges associated with their use as therapeutics: they are readily degraded by proteases, have poor membrane permeability and may have poor solubility. As a result, interest has shifted from naturally-occurring peptides towards modified peptides, peptidomimetics and cyclic peptide-based molecules.[14] Cyclic peptides are particularly promising as therapeutics, but they can be challenging to chemically synthesise, especially where the peptide is small and does not contain a turn-inducing element.[49,50]

Oxetane has previously been used as an isosteric replacement for carbonyl groups in medicinal chemistry, as the hydrogen bonding capabilities and lone pair electron arrangement are similar between the two chemical groups.[53] For example, Carreira and colleagues performed carbonyl to oxetane substitution on the drugs thalidomide and lenalidomide. Modification of thalidomide with an oxetane heterocycle blocks the *in vivo* racemisation, which could prevent the teratogenic effects of the (−)-*S* enantiomer.[60]

Recently, Shipman and colleagues developed oxetane-modified peptidomimetics, and demonstrated that the termini of the modified peptides were closer together more frequently, suggesting that oxetane modification may assist in the preparation of cyclic peptide-based molecules.[7] Much of the work presented here involved these oxetane-modified peptides, but the effect of azetidine modification on small peptides was also discussed. Like oxetane, azetidine is a four-membered heterocycle, but with an NH in place of the oxygen. Azetidines have similarly

been used in medicinal chemistry since the 1970s to enhance the pharmacokinetic properties of small molecule drugs,[66–69] and more recently have been attracting attention due to their ability to be selectively functionalised.[75–77] Additionally, the azetidine ring is much more stable than oxetane under the acidic conditions required to remove protecting groups from certain amino acid sidechains such as arginine.[41,53]

This study aimed to characterise the effect of oxetane and azetidine modification on small peptides using a combination of biophysical, microbiological and computational techniques. In order to achieve this aim, the study was divided into three smaller objectives:

- Characterise the impact of oxetane modification on small peptides (cyclic and linear) using a combination of biophysical and computational techniques.
- Characterise the impact of oxetane modification on longer peptides with α-helical content (an important motif in biology), using a predominantly computational approach.
- Assess changes in antimicrobial activity following cyclisation and/or azetidine modification for a model sequence with known biological activity.

7.2 Key Findings

In this work, we have demonstrated that oxetane modification of a pentapeptide LAGAY induces formation of a turn, as evidenced by the presence of medium-range NOEs which are absent in the unmodified parent peptide. This was the first experimental evidence to suggest that oxetane modification induces a turn in linear peptides. MD simulations with NMR-derived restraints for both LAGAY-OMe and $LAG_{ox}AY$-OMe suggest that $LAG_{ox}AY$-OMe is much more flexible than the unmodified peptide, spending over 10% of its time in a conformation where the termini are within 5 Å. This is in good agreement with previous work by Powell et al., in which oxetane-modified tripeptides were more flexible than their unmodified counterparts.[7] This also agrees with experimental evidence showing that oxetane-modified peptides have an increased yield of macrocyclisation.[41,92] A similar effect was observed in the same sequence following azetidine modification, and 2D NMR techniques reveal a similar pattern of medium-range NOEs. It was not possible to perform the corresponding MD simulations of this system as CHARMM-compatible force field parameters for the azetidine modification have not been developed at the time of writing.

The effect of oxetane modification on a cyclic pentapeptide cLAGAY was also assessed. CD spectroscopy revealed clear structural differences between modified and unmodified peptides, which were further probed using NMR techniques. There were notable differences in the NOE patterns between the parent and oxetane-modified cyclic peptides. MD simulations with NMR-derived restraints suggest that oxetane rigidified this peptide via the formation of an additional intramolecular hydrogen bond; however this effect may be sequence-specific. These simulations also that highlighted some improvements could be made to the force field parameters for oxetane, as there was a significant restraint violation in the cyclic modified peptide.

It is important to characterise the effect of any backbone modification on secondary structural motifs, which have vital roles in a multitude of biological processes. In this work, we explored the impact of oxetane modification on well-characterised helical peptides. CD spectra showed that there was a drastic reduction in helicity following oxetane modification, which was most pronounced when the modification was towards the middle of the sequence.[9,93] NMR data also revealed chemical shift changes in a modified peptide that corresponded with a loss of helicity compared to the unmodified parent.

Steered MD (SMD) was used to gain insights into the molecular determinants for the reduction in helicity. SMD can also provide insight into which regions of the peptide are unwind more readily, which cannot be obtained using techniques such as CD. Changes to the dihedral angles and hydrogen bonding patterns in proximity to the modification were observed, and this disruption to the helical structure was shown to reduce the work needed to unwind the helices. Unwinding the peptides using SMD subjects the peptides to forces that do not occur under normal circumstances, therefore it is vital to validate any results by comparing to experimental data. These simulations did not include any NMR-derived restraints, so it is remarkable how well the results from the SMD match the trend of global helicity observed using CD.

Although one might consider a carbonyl to an oxetane substitution a relatively conservative modification, this work demonstrates that any impact on dihedral angles and hydrogen bonding must be considered when modifying a helical structure. Together the data suggest that oxetane is not a useful modification for stabilising helical structures, but instead may be better suited in other structures where they may stabilise or mimic β-turn structural motifs. Interestingly, oxetane-modified glycine is less disruptive to a helical peptide than oxetane-modified alanine, likely because under normal conditions alanine is considered stabilising to helices whilst glycine is disruptive.[192] Oxetane modification appears to disrupt alanine's ability to stabilise helices and, to a lesser degree, glycine's ability to disrupt them.

The effect of azetidine modification on an AMP was also explored. Due to the relative instability of the oxetane ring under acidic conditions, it was not possible to prepare the equivalent oxetane-modified AMP.[41] Initially, bacaucin-1 and bacaucin-1a were selected as model systems, but we were unable to reproduce the published biological activity for these peptides, and the authors of the original publications could not reproduce their published data either. Instead, a panel of five AMPs with published activity were screened against *E. coli* and *S. aureus*. The most promising target, Ac-RRWWCA-NH_2, was selected for modification and/or cyclisation. Firstly, the cysteine residue was modified to a glycine, which rendered the peptide inactive, even following cyclisation. However, azetidine modification on the glycine restored antimicrobial activity against *E. coli*, without increasing the haemolytic activity against mammalian erythrocytes. Cyclisation of Ac-RRWWCA-NH_2 enhanced the antimicrobial activity of the peptide, in line with previous work which demonstrated that cyclisation often increases the activity of AMPs.[90,91] Unexpectedly, the linear azetidine-modified AMP demonstrated higher biological activity compared to the cyclic azetidine-modified variant, with improved activity compared to the original sequence of Ac-RRWWCA-NH_2. This activity is bacteriostatic rather than bactericidal, suggesting that the AMPs act on an intracellular target, rather than at the bacterial membrane.

Together, the work presented here shows that oxetane is a useful modification in the development of cyclic peptidomimetics, due to its turn-inducing effect. However, this same effect renders it unsuitable for use in helical motifs. Oxetane modification remains limited by the residues at which the modification can be incorporated, as well as its relative instability under acidic conditions (it is prone to ring-opening).[53] Azetidines are much more stable under the acidic conditions required to remove sidechain protecting groups,[41] and as a result azetidine modification remains a promising area moving forward. In addition, azetidines have the ability to be selectively functionalised. This work shows that it is possible to restore activity of an AMP through azetidine modification. Overall, the results shown here may help direct future design of chemically modified peptides.

7.3 Future Work

The effect of oxetane and azetidine modification of small peptides has begun to be explored, but there are many areas that can continue to be investigated.

In Chapter 4, the impact of oxetane modification on a cyclic pentapeptide was explored. We noted some changes in the amide region of the ^{1}H NMR spectra following dilution for both cLAGAY and cLAG_{ox}AY. Such chemical shift pertur-

bations could be indicative of intermolecular interactions such as oligomerisation. This should be further investigated, although this is likely to be challenging due to the small size of these peptides. Mass spectrometry may be able to provide information about their oligomeric state. Cyclic peptides have previously been described to self-assemble into hollow nanotubes, although this appears to occur predominantly in sequences with β- or γ-amino acids, or alternating L- and D-α-amino acids.[221–224] As the sequences we studied contained only L-α-amino acids, it is unclear whether they would be capable of such large-scale self-assembly.

The effect of oxetane modification on a well-characterised α-helical structure was assessed using steered MD (Chapter 5). This work could be complemented with a similar study comparing oxetane-modified and parent β-sheet structures. Indeed, preliminary work by Kitchner suggests that oxetane modification at the turn of a β-hairpin may have a stabilising effect, while modification on the strand of the hairpin destabilises the structure.[225] Understanding the effect of oxetane modification on secondary structure motifs has important implications in the future design of peptide-based therapeutics.

In this work, we assessed the impact of azetidine modification on antimicrobial activity (Chapter 6). It would be valuable to characterise how azetidine modification changes the structure of small peptides. However, it was not possible to generate MD structures of azetidine-modified peptides for this study as at the time of writing the force field parameters to correctly represent this chemistry do not exist. Therefore, parameterisation would allow for molecular modelling techniques to be applied to azetidine-modified peptides in a manner similar to the oxetane-modified peptides. Some structural information surrounding the azetidine modification is known: as described in this work, azetidine-modified LAGAY demonstrates NOEs indicative of turn formation, and Saunders[41] was able to produce a crystal structure of a cyclic tetrapeptide cVG_{az}($COCH_2N_3$)LW(Boc) (Fig. 7.1). Unfortunately, it was not possible to obtain crystal structures of the unmodified tetrapeptide; therefore it would be valuable to obtain solution structures of both the unmodified and modified structures using NMR.

Comparison of the crystal structure of cVG_{az}($COCH_2N_3$)LW(Boc) to the MD structures of cLAG_{ox}AY reveals key differences between azetidine and oxetane modification. For example, the ω angle of cLAG_{ox}AY is 69.49°, whilst the ω angle of cVG_{az}($COCH_2N_3$)LW(Boc) is 158.72°, much closer to the expected angle for a *trans* amide bond.[41] However, these changes may be due to differences in the ring size—pentapeptides are less strained than tetrapeptides, allowing for more conformational flexibility surrounding the oxetane ring. The two peptides also have different amino acid sidechains. Therefore, it would be useful to generate NMR-restrained MD structures for LAG_{az}AY-OMe and cLAG_{az}AY, which could

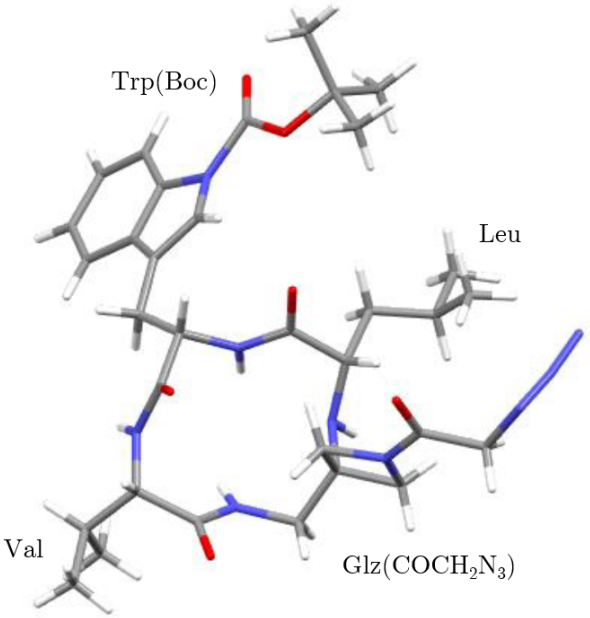


Figure 7.1: Mercury plot of the crystal structure of $cVG_{az}(COCH_2N_3)LW(Boc)$. Reproduced from Saunders.[41]

be directly compared to the unmodified parent and to oxetane-modified LAGAY.

There is also the scope to investigate the effect of cysteine in the antimicrobial sequence Ac-RRWWCA-NH_2, and the mechanism of action by which this peptide exerts its antimicrobial activity. In particular, the inconsistency of the activity of cRRWWCA between repeats following addition of DTT could be explored further. This compound appears to be changing over time, as there were differences between the preparative and analytical HPLC. Heteronuclear NMR techniques can be used to determine the redox state of cysteine residues, by using the $^{13}C\beta$ chemical shift data as a sensitive measure,[226] so this could provide some insight into whether the peptide is forming intermolecular disulphide bonds in the absence of DTT.

Small molecule azetidine derivatives have previously been demonstrated to exert promising antimicrobial activities,[210,211] but the mechanisms by which azetidine increases antimicrobial activity of a compound remains unclear. It would therefore be prudent to explore the mechanism by which azetidine-modified AMPs act upon bacteria using techniques such as scanning electron microscopy (SEM). SEM has previously been used to monitor changes in cell morphology following treatment with AMPs.[227,228] Additionally, it would be useful to screen the unmodified sequence RWWGAR, as this could be directly compared to the azetidine-modified sequence which demonstrated the most activity, $RWWG_{az}AR$.

Perhaps the most significant area that can be improved upon for the development of peptide-based therapeutics is the limited residues which can be modified with oxetane or azetidine (the synthetic chemistry techniques used are currently limited to glycine or alanine modifications). This was a major limitation when selecting a sequence to screen for antimicrobial activity, as the selected sequence

ideally needed to contain a glycine or alanine (or a residue that could be replaced with one of these residues), and needed to be less than seven amino acids in length, as sequences longer than this are significantly easier to cyclise using conventional methods,[49,50] thus the azetidine modification is less advantageous. By increasing the number of amino acids that can be modified with oxetanes or azetidines, the scope would be drastically increased. For example, many antimicrobial peptides are highly cationic, often containing predominantly arginine and tryptophan residues, with very few aliphatic residues. If tryptophan or arginine residues were able to be oxetane- or azetidine-modified, it would be possible to study the antimicrobial activity of a more conventional AMP with and without a modification.

Bibliography

[1] A. F. Räder, M. Weinmüller, F. Reichart, A. Schumacher-Klinger, S. Merzbach, C. Gilon, A. Hoffman and H. Kessler, *Angewandte Chemie - International Edition*, 2018, **57**, 14414–14438.

[2] L. Gentilucci, R. De Marco and L. Cerisoli, *Current Pharmaceutical Design*, 2010, **16**, 3185–3203.

[3] A. Zega, *Current Medicinal Chemistry*, 2005, **12**, 589–597.

[4] R. N. Zuckermann and T. Kodadek, *Current Opinion in Molecular Therapeutics*, 2009, **11**, 299–307.

[5] A. A. Vinogradov, Y. Yin and H. Suga, *Journal of the American Chemical Society*, 2019, **141**, 4167–4181.

[6] G. M. Pauletti, S. Gangwar, T. J. Siahaan, J. Aubé and R. T. Borchardt, *Advanced Drug Delivery Reviews*, 1997, **27**, 235–256.

[7] N. H. Powell, G. J. Clarkson, R. Notman, P. Raubo, N. G. Martin and M. Shipman, *Chemical Communications*, 2014, **50**, 8797.

[8] J. D. Beadle, A. Knuhtsen, A. Hoose, P. Raubo, A. G. Jamieson and M. Shipman, *Organic Letters*, 2017, **19**, 3303–3306.

[9] J. D. Beadle, *Ph.D. thesis*, University of Warwick, 2018.

[10] S. Roesner, J. D. Beadle, L. K. B. Tam, I. Wilkening, G. J. Clarkson, P. Raubo and M. Shipman, *Organic and Biomolecular Chemistry*, 2020, **18**, 5400–5405.

[11] C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Advanced Drug Delivery Reviews*, 1997, **23**, 3–25.

[12] C. A. Lipinski, *Journal of Pharmacological and Toxicological Methods*, 2000, **44**, 235–249.

[13] G. Li, K. Shao and C. S. Umeshappa, in *Brain Targeted Drug Delivery System*, ed. H. Gao and X. Gao, Academic Press, 2019, pp. 33–51.

[14] B. J. Bruno, G. D. Miller and C. S. Lim, *Therapeutic Delivery*, 2014, **4**, 1443–1467.

[15] J. A. Lichty, *Transactions of the American Clinical and Climatological Association*, 1923, **39**, 119–134.

[16] D. J. Craik, D. P. Fairlie, S. Liras and D. Price, *Chemical Biology and Drug Design*, 2013, **81**, 136–147.

[17] H. X. Ngo and S. Garneau-Tsodikova, *MedChemComm*, 2018, **9**, 757–758.

[18] T. Morrow and L. H. Felcone, *Biotechnology Healthcare*, 2004, **1**, 24–29.

[19] A. Chakrabartty, T. Kortemme and R. L. Baldwin, *Protein Science*, 1994, **3**, 843–852.

[20] A. Henninot, J. C. Collins and J. M. Nuss, *Journal of Medicinal Chemistry*, 2018, **61**, 1382–1414.

[21] K. Fosgerau and T. Hoffmann, *Drug Discovery Today*, 2015, **20**, 122–128.

[22] S. Lee, J. Xie and X. Chen, *Chemical Reviews*, 2010, **110**, 3087–3111.

[23] A. Dasgupta, J. H. Mondal and D. Das, *RSC Advances*, 2013, **3**, 9117–9149.

[24] A. Loffet, *Journal of Peptide Science*, 2002, **8**, 1–7.

[25] A. Isidro-Llobet, M. N. Kenworthy, S. Mukherjee, M. E. Kopach, K. Wegner, F. Gallou, A. G. Smith and F. Roschangar, *Journal of Organic Chemistry*, 2019, **84**, 4615–4628.

[26] J. Renukuntla, A. D. Vadlapudi, A. Patel, S. H. Boddu and A. K. Mitra, *International Journal of Phamaceutics*, 2014, **447**, 4–6.

[27] L. Sun, *Modern Chemistry and Applications*, 2013, **1**, 1–2.

[28] K. Ito, T. Passioura and H. Suga, *Molecules*, 2013, **18**, 3502–3528.

[29] C. K. Wang and D. J. Craik, *Biopolymers*, 2016, **106**, 901–909.

[30] M. Erak, K. Bellmann-Sickert, S. Els-Heindl and A. G. Beck-Sickinger, *Bioorganic and Medicinal Chemistry*, 2018, **26**, 2759–2765.

[31] H. R. Jensen, M. F. Laursen, D. L. Lildballe, J. B. Andersen, E. Nexø and T. R. Licht, *BMC Research Notes*, 2011, **4**, 208.

[32] A. K. Petrus, T. J. Fairchild and R. P. Doyle, *Angewandte Chemie - International Edition*, 2009, **48**, 1022–1028.

[33] S. Chen, D. Gfeller, S. A. Buth, O. Michielin, P. G. Leiman and C. Heinis, *ChemBioChem*, 2013, **14**, 1316–1322.

[34] Z. Feng and B. Xu, *Biomolecular Concepts*, 2016, **7**, 179–187.

[35] X. Wei, C. Zhan, Q. Shen, W. Fu, C. Xie, J. Gao, C. Peng, P. Zheng and W. Lu, *Angewandte Chemie - International Edition*, 2015, **54**, 3023–3027.

[36] *Peptide-based drug design: here and now*, ed. L. Otvos, Humana Press, 2008.

[37] M. Morishita and N. A. Peppas, *Drug Discovery Today*, 2006, **11**, 905–910.

[38] E. Lenci and A. Trabocchi, *Chemical Society Reviews*, 2020, **49**, 3262–3277.

[39] I. Avan, C. Dennis Hall and A. R. Katritzky, *Chemical Society Reviews*, 2014, **43**, 3575–3594.

[40] J. Deska and U. Kazmaier, *Current Organic Chemistry*, 2008, **12**, 355–385.

[41] G. J. Saunders, *Ph.D. thesis*, University of Warwick, 2020.

[42] M. A. Abdalla and L. J. McGaw, *Molecules*, 2018, **23**, 1–19.

[43] C. J. White and A. K. Yudin, *Nature Chemistry*, 2011, **3**, 509–524.

[44] S. H. Joo, *Biomolecules and Therapeutics*, 2012, **20**, 19–26.

[45] A. F. Räder, F. Reichart, M. Weinmüller and H. Kessler, *Bioorganic and Medicinal Chemistry*, 2018, **26**, 2766–2773.

[46] A. Tapeinou, M. T. Matsoukas, C. Simal and T. Tselios, *Biopolymers*, 2015, **104**, 453–461.

[47] M. Aumailley, M. Gurrath, G. Müller, J. Calvete, R. Timpl and H. Kessler, *FEBS Letters*, 1991, **29**, 50–54.

[48] W. Bauer, U. Briner, W. Doepfner, R. Haller, R. Huguenin, P. Marbach, T. J. Petcher and J. Pless, *Life Sciences*, 1982, **31**, 1133–1140.

[49] E. N. Prabhakaran, I. N. Rao, A. Boruah and J. Iqbal, *Journal of Organic Chemistry*, 2002, **67**, 8247–8250.

[50] P. R. Reddy, V. Balraju, G. R. Madhavan, B. Banerji and J. Iqbal, *Tetrahedron Letters*, 2003, **44**, 353–356.

[51] J. A. Burkhard, G. Wuitschik, M. Rogers-Evans, K. Müller and E. M. Carreira, *Angewandte Chemie - International Edition*, 2010, **49**, 9052–9067.

[52] C. M. Marson, *Chemical Society Reviews*, 2011, **40**, 5514–5533.

[53] J. A. Bull, R. A. Croft, O. A. Davis, R. Doran and K. F. Morgan, *Chemical Reviews*, 2016, **116**, 12150–12233.

[54] M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon and A. T. Mcphail, *Journal of the American Chemical Society*, 1971, **93**, 2325–2327.

[55] N. Shimada, S. Hasegawa, T. Harada, T. Tomisawa, A. Fujii and T. Takita, *Journal of Antibiotics*, 1986, **39**, 1623.

[56] S. Ōmura, M. Murata, N. Imamura, Y. Iwai, H. Tanaka, A. Furusaki and T. Matsumoto, *Journal of Antibiotics*, 1984, **37**, 1324–1332.

[57] G. Wuitschik, E. M. Carreira, B. Wagner, H. Fischer, I. Parrilla, F. Schuler, M. Rogers-Evans and K. Müller, *Journal of Medicinal Chemistry*, 2010, **53**, 3227–3246.

[58] G. Wuitschik, M. Rogers-Evans, K. Müller, H. Fischer, B. Wagner, F. Schuler, L. Polonchuk and E. M. Carreira, *Angewandte Chemie - International Edition*, 2006, **45**, 7736–7739.

[59] P. Lassalas, K. Oukoloff, V. Makani, M. James, V. Tran, Y. Yao, L. Huang, K. Vijayendran, L. Monti, J. Q. Trojanowski, V. M. Lee, M. C. Kozlowski, A. B. Smith, K. R. Brunden and C. Ballatore, *ACS Medicinal Chemistry Letters*, 2017, **8**, 864–868.

[60] J. A. Burkhard, G. Wuitschik, J. M. Plancher, M. Rogers-Evans and E. M. Carreira, *Organic Letters*, 2013, **15**, 4312–4315.

[61] S. J. Matthews and C. McCoy, *Clinical Therapeutics*, 2003, **25**, 342–395.

[62] M. McLaughlin, R. Yazaki, T. C. Fessard and E. M. Carreira, *Organic Letters*, 2014, **16**, 4070–4073.

[63] G. P. Möller, S. Müller, B. T. Wolfstädter, S. Wolfrum, D. Schepmann, B. Wünsch and E. M. Carreira, *Organic Letters*, 2017, **19**, 2510–2513.

[64] A. Isidro-Llobet, M. Álvarez and F. Albericio, *Chemical Reviews*, 2009, **109**, 2455–2504.

[65] A. Žukauskaite, A. Moretto, C. Peggion, M. De Zotti, A. Šačkus, F. Formaggio, N. De Kimpe and S. Mangelinckx, *European Journal of Organic Chemistry*, 2014, **2014**, 2312–2321.

[66] P. Melloni, A. Della Torre, M. Meroni, A. Ambrosini and A. C. Rossi, *Journal of Medicinal Chemistry*, 1979, **2**, 183–191.

[67] J. A. Burkhard, B. Wagner, H. Fischer, F. Schuler, K. Müller and E. M. Carreira, *Angewandte Chemie - International Edition*, 2010, **49**, 3524–3527.

[68] D. J. St. Jean and C. Fotsch, *Journal of Medicinal Chemistry*, 2012, **55**, 6002–6020.

[69] J. Zhou, X. Lin, X. Ji, S. Xu, C. Liu, X. Dong, W. Zhao and W. Zhao, *Organic Letters*, 2020, **22**, 4413–4417.

[70] Z. T. Al-Salama and L. J. Scott, *Drugs*, 2018, **78**, 761–772.

[71] K. P. Garnock-Jones, *Drugs*, 2015, **75**, 1823–1830.

[72] M. Kitajima, N. Kogure, K. Yamaguchi, H. Takayama and N. Aimi, *Organic Letters*, 2003, **5**, 2075–2078.

[73] R. Gianatassio, J. M. Lopchuk, J. Wang, C. M. Pan, L. R. Malins, L. Prieto, T. A. Brandt, M. R. Collins, G. M. Gallego, N. W. Sach and J. E. Spangler, *Science*, 2016, **6270**, 241–246.

[74] J. A. Milligan and P. Wipf, *Nature Chemistry*, 2016, **4**, 296.

[75] N. D. Barth, R. Subiros-Funosas, L. Mendive-Tapia, R. Duffin, M. A. Shields, J. A. Cartwright, S. T. Henriques, J. Sot, F. M. Goñi, R. Lavilla, J. A. Marwick, S. Vermeren, A. G. Rossi, M. Egeblad, I. Dransfield and M. Vendrell, *Nature Communications*, 2020, **11**, 1–14.

[76] E. Ragozin, A. Hesin, A. Bazylevich, H. Tuchinsky, A. Bovina, T. Shekhter Zahavi, M. Oron-Herman, G. Kostenich, M. A. Firer, T. Rubinek, I. Wolf, G. Luboshits, M. Y. Sherman and G. Gellerman, *Bioorganic and Medicinal Chemistry*, 2018, **26**, 3825–3836.

[77] P. G. Dougherty, A. Sahni and D. Pei, *Chemical Reviews*, 2019, **119**, 10241–10287.

[78] M. S. Mulani, E. E. Kamble, S. N. Kumkar, M. S. Tawre and K. R. Pardesi, *Frontiers in Microbiology*, 2019, **10**, 1–24.

[79] Interagency Coordination Group on Antimicrobial Resistance, *No Time to Wait: Securing the future from drug-resistant infections*, World Health Organisation technical report, 2019.

[80] A. Pfalzgraff, K. Brandenburg and G. Weindl, *Frontiers in Pharmacology*, 2018, **9**, 1–23.

[81] L. J. Zhang and R. L. Gallo, *Current Biology*, 2016, **26**, R14–R19.

[82] K. V. Reddy, R. D. Yedery and C. Aranha, *International Journal of Antimicrobial Agents*, 2004, **24**, 536–547.

[83] L. B. Oyama, S. E. Girdwood, A. R. Cookson, N. Fernandez-Fuentes, F. Privé, H. E. Vallin, T. J. Wilkinson, P. N. Golyshin, O. V. Golyshina, R. Mikut, K. Hilpert, J. Richards, M. Wootton, J. E. Edwards, M. Maresca, J. Perrier, F. T. Lundy, Y. Luo, M. Zhou, M. Hess, H. C. Mantovani, C. J. Creevey and S. A. Huws, *npj Biofilms and Microbiomes*, 2017, **3**, 1–9.

[84] T. T. Thomsen, H. C. Mendel, W. Al-mansour, A. Oddo, A. Løbner-olesen and P. R. Hansen, *Antibiotics*, 2020, **9**, 1–13.

[85] L. T. Nguyen, E. F. Haney and H. J. Vogel, *Trends in Biotechnology*, 2011, **29**, 464–472.

[86] D. I. Chan, E. J. Prenner and H. J. Vogel, *Biochimica et Biophysica Acta - Biomembranes*, 2006, **1758**, 1184–1202.

[87] X. Bi, C. Wang, L. Ma, Y. Sun and D. Shang, *Journal of Applied Microbiology*, 2013, **115**, 663–672.

[88] J. C. Ma and D. A. Dougherty, *Chemical Reviews*, 1997, **97**, 1303–1324.

[89] A. Falanga, E. Nigro, M. G. De Biasi, A. Daniele, G. Morelli, S. Galdiero and O. Scudiero, *Molecules*, 2017, **22**, 1–15.

[90] J. T. Mika, G. Moiset, A. D. Cirac, L. Feliu, E. Bardají, M. Planas, D. Sengupta, S. J. Marrink and B. Poolman, *Biochimica et Biophysica Acta - Biomembranes*, 2011, **1808**, 2197–2205.

[91] M. Dathe, H. Nikolenko, J. Klose and M. Bienert, *Biochemistry*, 2004, **43**, 9140–9150.

[92] S. Roesner, G. Saunders, I. Wilkening, E. Jayawant, J. Geden, P. Kerby, A. Dixon, R. Notman and M. Shipman, *Chemical Science*, 2019, **10**, 2465–2472.

[93] E. S. Jayawant, J. D. Beadle, I. Wilkening, P. Raubo, M. Shipman, R. Notman and A. M. Dixon, *Physical Chemistry Chemical Physics*, 2020, **22**, 25075–25083.

[94] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten, *Structure*, 2009, **17**, 1295–1306.

[95] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications (Vol. 1)*, Academic Press, San Diego, 2nd edn., 2001.

[96] A. R. Leach, *Molecular modelling: principles and applications*, Harlow: Prentice Hall, 2001.

[97] A. D. Mackerell Jr., N. Banavali and N. Foloppe, *Biopolymers: Original Research on Biomolecules*, 2001, **56**, 257–265.

[98] X. Zhu, P. E. Lopes and A. D. Mackerell Jr., *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 167–185.

[99] B. R. Brooks, C. L. Brooks, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545–1614.

[100] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.

[101] P. Ewald, *Ann. Phys.*, 1921, **64**, 253–287.

[102] J. A. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, Springer US, 2005.

[103] M. R. Hestenes and E. Stiefel, *Journal of Research of the National Bureau of Standards*, 1952, **49**, 409–436.

[104] R. Fletcher, *Computer Journal*, 1970, **13**, 317–322.

[105] C. G. Broyden, *IMA Journal of Applied Mathematics*, 1970, **6**, 76–90.

[106] C. G. Broyden, *IMA Journal of Applied Mathematics*, 1970, **6**, 222–231.

[107] T. Hansson, C. Oostenbrink and W. F. Van Gunsteren, *Current opinion in structural biology*, 2002, **12**, 190–196.

[108] M. Karplus and G. A. Petsko, *Nature*, 1990, **347**, 631–639.

[109] W. F. Van Gunsteren and H. J. Berendsen, *Molecular Simulation*, 1988, **1**, 173–185.

[110] L. Verlet, *Physical Review*, 1967, **159**, 98–103.

[111] J. Willard Gibbs, *Elementary Principles in Statistical Mechanics*, Charles Scribner's Sons, New York, 1902.

[112] H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.

[113] S. Nosé and M. L. Klein, *Molecular Physics*, 1983, **50**, 1055–1076.

[114] G. Bussi, D. Donadio and M. Parrinello, *Journal of Chemical Physics*, 2007, **126**, 014101.

[115] M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.

[116] B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *Journal of Computational Chemistry*, 1997, **18**, 1463–1472.

[117] B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435–447.

[118] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers and K. Schulten, in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, ed. P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich and R. D. Skeel, Springer-Verlag, Berlin, 1st edn., 1999, ch. Steered Mo, pp. 39–65.

[119] B. Isralewitz, M. Gao and K. Schulten, *Current Opinion in Structural Biology*, 2001, **11**, 224–230.

[120] J. A. Lemkul and D. R. Bevan, *Journal of Physical Chemistry B*, 2010, **114**, 100.

[121] C. A. López, A. H. de Vries and S. J. Marrink, *Carbohydrate Research*, 2012, **364**, 1–7.

[122] G. Ozer, S. Quirk and R. Hernandez, *Journal of Chemical Theory and Computation*, 2012, **8**, 4837–4844.

[123] M. Sotomayor, AIP Conference Proceedings, 2015.

[124] A. Rodger, in *Encyclopedia of Biophysics*, ed. G. C. K. Roberts, Springer, Berlin, Heidelberg, 2013, pp. 311–313.

[125] L. Whitmore and B. A. Wallace, *Biopolymers: Original Research on Biomolecules*, 2008, **89**, 392–400.

[126] T. D. Claridge, *High-resolution NMR techniques in organic chemistry*, Elsevier, 2016.

[127] A. D. Bax and G. D. Donald, *Journal of Magnetic Resonance*, 1985, **65**, 355–360.

[128] L. Braunschweiler and R. R. Ernst, *Journal of Magnetic Resonance*, 1983, **53**, 521–528.

[129] G. C. K. Roberts, *NMR of Macromolecules: A Practical Approach*, Oxford University Press, Oxford, 1993.

[130] P. Lundström, in *Encyclopedia of Biophysics*, ed. G. C. K. Roberts, Springer, Berlin, Heidelberg, 2013, pp. 1753–1759.

[131] M. P. Williamson, in *Modern Magnetic Resonance*, ed. G. A. Webb, Springer, Cham, 2017, ch. The Transf, pp. 1–15.

[132] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, 1986.

[133] C. Lockey, *Ph.D. thesis*, University of Warwick, 2017.

[134] T. L. Hwang and A. J. Shaka, *Journal of Magnetic Resonance - Series A*, 1995, **112**, 275–279.

[135] M. Liu, X.-A. Mao, C. Ye, H. Huang, J. K. Nicholson and J. C. Lindon, *Journal of Magnetic Resonance*, 1998, **132**, 125–129.

[136] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins: Structure, Function, and Bioinformatics*, 2005, **59**, 687–696.

[137] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.

[138] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess and E. Lindahl, *Journal of Chemical Theory and Computation*, 2010, **6**, 459–466.

[139] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchinson, *Journal of Cheminformatics*, 2012, **4**, 17.

[140] I. Vorobyov, V. M. Anisimov, S. Greene, R. M. Venable, A. Moser, R. W. Pastor and A. D. MacKerell Jr., *Journal of Chemical Theory and Computation*, 2007, **3**, 1120–1133.

[141] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid and J. C. Gumbart, *Journal of Computational Chemistry*, 2013, **34**, 2757–2770.

[142] W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

[143] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03*, Gaussian, Inc., Wallingford, CT, 2004.

[144] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *Journal of the American Chemical Society*, 1992, **114**, 10024–10035.

[145] M. L. Strader and S. E. Feller, *Journal of Physical Chemistry A*, 2002, **106**, 1074–1080.

[146] A. E. Torda, R. M. Scheek and W. F. van Gunsteren, *Chemical Physics Letters*, 1989, **157**, 289–294.

[147] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angewandte Chemie - International Edition*, 1999, **38**, 236–240.

[148] F. Jiang and H. Geng, in *Computational Methods for Studying Conformational Behaviors of Cyclic Peptides*, ed. G. Goetz, Springer New York, New York, NY, 2019, pp. 61–71.

[149] J. E. Sykes and S. C. Rankin, in *Canine and Feline Infectious Diseases*, Elsevier Inc., St. Louis, 2013, ch. 3, pp. 17–28.

[150] Y. Liu, M. Song, S. Ding and K. Zhu, *ACS Infectious Diseases*, 2019, **5**, 123–130.

[151] Y. Liu, S. Ding, R. Dietrich, E. Märtlbauer and K. Zhu, *Angewandte Chemie - International Edition*, 2017, **56**, 1486–1490.

[152] S. E. Blondelle, E. Takahashi, K. T. Dinh and R. A. Houghten, *Journal of Applied Bacteriology*, 1995, **78**, 39–46.

[153] D. Ifrah, X. Doisy, T. S. Ryge and P. R. Hansen, *Journal of Peptide Science*, 2005, **11**, 113–121.

[154] M. Pasupuleti, A. Schmidtchen, A. Chalupka, L. Ringstad and M. Malmsten, *PLoS ONE*, 2009, **4**, e5285.

[155] Y. Otsuka, T. Ishikawa, C. Takahashi and M. Masuda, *Toxins*, 2019, **11**, 392.

[156] J. Xiao, H. Zhang, L. Niu and X. Wang, *Journal of Agricultural and Food Chemistry*, 2011, **59**, 1145–1151.

[157] I. Wiegand, K. Hilpert and R. E. Hancock, *Nature Protocols*, 2008, **3**, 163–175.

[158] J. M. Andrews, *Journal of Antimicrobial Chemotherapy*, 2001, **48**, 5–16.

[159] S. Fernandez-Lopez, H. S. Kim, E. C. Choi, M. Delgado, J. R. Granja, A. Khasanov, K. Kraehenbuehl, G. Long, D. A. Weinberger, K. M. Wilcoxen and M. R. Ghadiri, *Nature*, 2001, **412**, 452–455.

[160] O. Boutureira, N. Martínez-Sáez, K. M. Brindle, A. A. Neves, F. Corzana and G. J. Bernardes, *Chemistry - A European Journal*, 2017, **23**, 6483–6489.

[161] N. Martínez-Sáez, S. Sun, D. Oldrini, P. Sormanni, O. Boutureira, F. Carboni, I. Compañón, M. J. Deery, M. Vendruscolo, F. Corzana, R. Adamo

and G. J. Bernardes, *Angewandte Chemie - International Edition*, 2017, **56**, 14963–14967.

[162] A. Bockus, C. McEwen and R. Lokey, *Current topics in medicinal chemistry*, 2013, **13**, 821–836.

[163] D. Nielsen, N. Shepherd, W. Xu, A. Lucke, M. Stoermer and D. Fairlie, *Chemical reviews*, 2017, **117**, 8094–8128.

[164] Y.-C. Tang, H.-B. Xie, G.-L. Tian and Y.-H. Ye, *Journal of Peptide Research*, 2002, **60**, 95–103.

[165] S. Hwang, Q. Shao, H. Williams, C. Hilty and Y. Q. Gao, *Journal of Physical Chemistry B*, 2011, **115**, 6653–6660.

[166] A. R. Gargaro, G. B. Bloomberg, C. E. Dempsey, M. Murray and M. J. Tanner, *European Journal of Biochemistry*, 1994, **221**, 445–454.

[167] S. M. Howell, S. V. Fiacco, T. T. Takahashi, F. Jalali-Yazdi, S. W. Millward, B. Hu, P. Wang and R. W. Roberts, *Scientific Reports*, 2014, **4**, 1–5.

[168] J. L. Kulp, J. C. Owrutsky, D. Y. Petrovykh, K. P. Fears, R. Lombardi, L. A. Nafie and T. D. Clark, *Biointerphases*, 2011, **6**, 1–7.

[169] N. Pirooznia, K. Abdi, D. Beiki, F. Emami, S. S. Arab, O. Sabzevari and S. Soltani-Gooshkhaneh, *Bioorganic Chemistry*, 2020, **102**, 104100.

[170] S. J. Bogdanowich-Knipp, D. S. Jois and T. J. Siahaan, *Journal of Peptide Research*, 1999, **53**, 523–529.

[171] Á. Roxin and G. Zheng, *Future Medicinal Chemistry*, 2012, **4**, 1601–1618.

[172] D. S. Nielsen, R. J. Lohman, H. N. Hoang, T. A. Hill, A. Jones, A. J. Lucke and D. P. Fairlie, *ChemBioChem*, 2015, **16**, 2289–2293.

[173] L. Pauling, R. B. Corey and H. R. Branson, *Proceedings of the National Academy of Sciences of the United States of America*, 1951, **37**, 205–211.

[174] A. P. Kornev, S. S. Taylor and L. F. Ten Eyck, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 14377–14382.

[175] M. Lemmon and J. Schlessinger, *Cell*, 2010, **141**, 1117–1134.

[176] R. L. Brooks and A. M. Dixon, *Biochimica et Biophysica Acta - Biomembranes*, 2020, **1862**, 183160.

[177] V. Azzarito, K. Long, N. S. Murphy and A. J. Wilson, *Nature Chemistry*, 2013, **5**, 161–173.

[178] A. L. Jochim and P. S. Arora, *Molecular BioSystems*, 2009, **5**, 924–926.

[179] I. Saraogi and A. D. Hamilton, *Biochemical Society Transactions*, 2008, **36**, 1414–1417.

[180] R. W. Cheloha, A. Maeda, T. Dean, T. J. Gardella and S. H. Gellman, *Nature Biotechnology*, 2014, **32**, 653–655.

[181] R. Banerjee, G. Basu, P. Chène and S. Roy, *Journal of Peptide Research*, 2002, **60**, 88–94.

[182] A. Chakrabartty, J. A. Schellman and R. L. Baldwin, *Nature*, 1991, **351**, 586–588.

[183] S. Marqusee, V. H. Robbins and R. L. Baldwin, *Proceedings of the National Academy of Sciences of the United States of America*, 1989, **86**, 5286–5290.

[184] A. Reiner, D. Wildemann, G. Fischer and T. Kiefhaber, *Journal of the American Chemical Society*, 2008, **130**, 8079–8084.

[185] D. S. Wishart, B. D. Sykes and F. M. Richards, *Biochemistry*, 1992, **31**, 1647–1651.

[186] R. B. Best, N.-V. Buchete and G. Hummer, *Biophysical Journal*, 2008, **95**, L07–L09.

[187] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell Jr., *Journal of Computational Chemistry*, 2010, **31**, 671–690.

[188] E. M. Meiering, *PLoS Biology*, 2008, **6**, 1383–1385.

[189] S. Walter, B. Hubner, U. Hahn and F. X. Schmid, *Journal of Molecular Biology*, 1995, **252**, 133–143.

[190] G. Desiraju and T. Steiner, *The weak hydrogen bond in structural chemistry and biology*, Oxford University Press/International Union of Crystallography, Oxford, 2001, pp. 1–480.

[191] N. Trinajstić, *Tetrahedron Letters*, 1968, **9**, 1529–1532.

[192] J. López-Llano, L. A. Campos, J. Sancho and J. Sancho, *Proteins: Structure, Function and Genetics*, 2006, **64**, 769–778.

[193] A. Grauer and B. König, *European Journal of Organic Chemistry*, 2009, 5099–5111.

[194] W. S. Horne, M. K. Yadav, C. D. Stout and M. R. Ghadiri, *Journal of the American Chemical Society*, 2004, **126**, 15366–15367.

[195] L. R. Whitby, Y. Ando, V. Setola, P. K. Vogt, B. L. Roth and D. L. Boger, *Journal of the American Chemical Society*, 2011, **133**, 10184–10194.

[196] S. Deike, S. Rothemund, B. Voigt, S. Samantray, B. Strodel and W. H. Binder, *Bioorganic Chemistry*, 2020, **101**, 104012.

[197] R. C. Founou, L. L. Founou and S. Y. Essack, *PLoS ONE*, 2017, **12**, 1–18.

[198] C. L. Ventola, *Pharmacy and Therapeutics*, 2015, **40**, 277–283.

[199] A. Fleming, *Bulletin of the World Health Organization*, 1929, **79**, 780–790.

[200] ReAct, *Tracking antimicrobial resistance in the Sustainable Development Goals*, `https://www.reactgroup.org/news-and-views/news-and-opinions/year-2019/tracking-antimicrobial-resistance-in-the-sustainable-development-goals/`, 2019, Accessed 11th March, 2021.

[201] A. Malina and Y. Shai, *Biochemical Journal*, 2005, **390**, 695–702.

[202] R. Notman, M. Noro, B. O'Malley and J. Anwar, *Journal of the American Chemical Society*, 2006, **128**, 13982–13983.

[203] H. I. Zgurskaya, G. Krishnamoorthy, A. Ntreh and S. Lu, *Frontiers in Microbiology*, 2011, **2**, 1–13.

[204] K. Andreev, M. W. Martynowycz, A. Ivankin, M. L. Huang, I. Kuzmenko, M. Meron, B. Lin, K. Kirshenbaum and D. Gidalevitz, *Langmuir*, 2016, **32**, 12905–12913.

[205] A. D. Cirac, G. Moiset, J. T. Mika, A. Koçer, P. Salvador, B. Poolman, S. J. Marrink and D. Sengupta, *Biophysical Journal*, 2011, **100**, 2422–2431.

[206] H. L. Chen, P. Y. Su, S. C. Kuo, T. L. Y. Lauderdale and C. Shih, *Frontiers in Microbiology*, 2018, **9**, 1–9.

[207] S. Srivastava, K. Dashora, K. L. Ameta, N. P. Singh, H. A. El-Enshasy, M. C. Pagano, A. E. L. Hesham, G. D. Sharma, M. Sharma and A. Bhargava, *Phytotherapy Research*, 2021, **35**, 256–277.

[208] P. Baindara, A. Kapoor, S. Korpole and V. Grover, *World Journal of Microbiology and Biotechnology*, 2017, **33**, 1–7.

[209] G. A. Pankey and L. D. Sabath, *Clinical Infectious Diseases*, 2004, **38**, 864–870.

[210] M. Asif, *MOJ Bioorganic & Organic Chemistry*, 2018, **2**, 36–40.

[211] R. S. Keri, K. M. Hosamani, H. S. Reddy and R. V. Shingalapur, *Archiv der Pharmazie - Chemistry in Life Sciences*, 2010, **343**, 237–247.

[212] J. J. Jackson and H. Kropp, *Journal of Infectious Diseases*, 1992, **165**, 1033–1041.

[213] V. Quagliarello and W. M. Scheld, *The New England Journal of Medicine*, 1992, **327**, 864–872.

[214] M. M. Mustafa, J. Mertsola, O. Ramilo, X. Sâez-Llorens, R. C. Risser and G. H. McCracken, *Journal of Infectious Diseases*, 1989, **160**, 891–895.

[215] M. Arditi, K. R. Manogue, M. Caplan and R. Yogev, *Journal of Infectious Diseases*, 1990, **162**, 139–147.

[216] D. Jha, R. Mishra, S. Gottschalk, K.-H. Wiesmüller, K. Ugurbil, M. E. Maier and J. Engelmann, *Bioconjugate Chemistry*, 2011, **22**, 319–328.

[217] A. C. Fogaça, I. C. Almeida, M. N. Eberlin, A. S. Tanaka, P. Bulet and S. Daffre, *Peptides*, 2006, **27**, 667–674.

[218] E. S. Jayawant, J. Hutchinson, D. Gašparíková, C. Lockey, L. Pruñonosa Lara, C. Guy, R. L. Brooks and A. M. Dixon, *ChemBioChem*, 2021, **22**, 2430–2439.

[219] N. Raheem and S. K. Straus, *Frontiers in Microbiology*, 2019, **10**, 1–14.

[220] E. C. Spindler, J. D. Hale, T. H. Giddings, R. E. Hancock and R. T. Gill, *Antimicrobial Agents and Chemotherapy*, 2011, **55**, 1706–1716.

[221] M. Tarek, B. Maigret and C. Chipot, *Biophysical Journal*, 2003, **85**, 2287–2298.

[222] R. J. Brea, C. Reiriz and J. R. Granja, *Chemical Society Reviews*, 2010, **39**, 1448–1456.

[223] M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee and N. Khazanovich, *Nature*, 1993, **366**, 324–327.

[224] J. D. Hartgerink, T. D. Clark and M. R. Ghadiri, *Chemistry - A European Journal*, 1998, **4**, 1367–1372.

[225] A. Kitcher, *M.Sc. thesis*, University of Warwick, 2019.

[226] C. Wiedemann, A. Kumar, A. Lang and O. Ohlenschläger, *Frontiers in Chemistry*, 2020, **8**, 1–8.

[227] K. G. Varnava, P. J. B. Edwards, E. Harjes, V. Sarojini and A. J. Cameron, *Journal of Peptide Science*, 2021, **27**, 1–11.

[228] M. Hartmann, M. Berditsch, J. Hawecker, M. F. Ardakani, D. Gerthsen and A. S. Ulrich, *Antimicrobial Agents and Chemotherapy*, 2010, **54**, 3132–3142.

Appendices

Script 7.1: Python 3 script to calculate area under curve (work done) via trapezium rule for force extension curves produced by steered molecular dynamics simulations.

```
from os import sys

delta = 0.000025
total_area = 0.0
fil = file(sys.argv[1],'r').readlines()
for i in range(0,len(fil)-1,2):
  h1 = float(fil[i].split()[1])
  h2 = float(fil[i+1].split()[1])
  area = delta*(h1+h2)/2.0
  total_area += area
print 'Total area under curve ', total_area
```